



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova

DIPLOMOVÁ PRÁCA

Bc. Soňa Šimonová

Rôzne metódy odhadu bodu zmeny

Katedra pravděpodobnosti a matematické statistiky

Vedúci diplomovej práce: doc. RNDr. Michal Pešta, PhD.

Študijný program: Matematika

Študijný odbor: Pravděpodobnost, matematická
statistika a ekonometrie

Praha 2020

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V dne

Podpis autora

Na tomto mieste by som sa chcela poďakovať doc. RNDr. Michalovi Peštovi, PhD. za vedenie tejto diplomovej práce a za cenné pripomienky a rady pri jej písaní.

Název práce: Rôzne metódy odhadu bodu zmeny

Autor: Bc. Soňa Šimonová

Katedra: Katedra pravdepodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Michal Pešta, PhD., Katedra pravdepodobnosti a matematické statistiky

Abstrakt: Cieľom tejto práce je podať zrozumiteľný výklad niektorých najnovších metód odhadovania bodu zmeny. V literatúre môžeme nájsť rôzne prístupy k riešeniu tejto problematiky. Medzi ne patria testy založené na procese CUSUM, testy pomerom vierohodností, metóda WBS a metódy pre odhadovanie bodu zmeny v panelových dátach. Okrem spomínaných prístupov je v práci vysvetlený aj prípad závislých panelov. Praktická časť štúdie je zameraná na aplikáciu metódy WBS na týždenné logaritmické výnosy akciového indexu Dow Jones. Najskôr na analyzovaný časový rad aplikujeme GARCH model. Metódu WBS následne použijeme na detekciu štrukturálnych zmien v pôvodnom rade, ako aj v rezíduách z modelu GARCH. Analyzujeme niekoľko typov penalizačných kritérií a skúmame ich vplyv na odhadnutý počet a polohy bodov zmeny v danom dátovom súbore.

Klíčová slova: bod zmeny, CUSUM štatistika, pomer vierohodností, binárna segmentácia, panelové dáta, okrajová podmienka, bootstrap, závislé dáta

Title: Various change point estimation methods

Author: Bc. Soňa Šimonová

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Michal Pešta, PhD., Department of Probability and Mathematical Statistics

Abstract: This thesis aims to give a comprehensive account of some of the most recent methods of a change point estimation. The literature on the change point estimation shows a variety of approaches to deal with this subject. Among them, tests based on the popular CUSUM process, likelihood ratio tests, wild binary segmentation and some of the most recent techniques on the change point estimation in panel data are all covered by this paper. The case of dependent panels is discussed as well. The practical part of the study is focused on application of the wild binary segmentation method on weekly log-returns of the Dow Jones stock index. Firstly, we fit a GARCH model to the analysed time series. We next use the wild binary segmentation method to detect structural changes in the mean of the original time series. Next, we apply the same method to the residuals from the GARCH fit. We analyse several penalization criteria proposed by previous studies and evaluate their effects on the estimated number and locations of the change points in the given data set.

Keywords: change point, CUSUM statistic, likelihood ratio, binary segmentation, panel data, boundary issue, bootstrap, dependent data

Obsah

Úvod	3
1 Štruktúrálné zmeny v časových radoch	4
1.1 Model časového radu s bodom zmeny	4
1.2 CUSUM procedúry	4
1.2.1 Štruktúrálné zmeny v nepodmienenej strednej hodnote . . .	5
1.3 Testy pomerom vierohodností	6
1.4 Odhad počtu bodov zmeny	7
2 Binárna segmentácia a WBS metóda	8
2.1 Model a predpoklady	8
2.2 Štandardná binárna segmentácia	9
2.3 WBS metóda	10
2.3.1 Zosilnené Schwarzovo informačné kritérium pre WBS	12
3 Odhady bodov zmeny pre panelové dáta	14
3.1 Odhad bodu zmeny metódou najmenších štvorcov	14
3.1.1 Limitné rozdelenie	16
3.2 Odhad bodu zmeny bez okrajovej podmienky	19
3.2.1 Odhad pre závislé dáta	23
3.3 Odhad bodu zmeny pomocou kvazi-vierohodnostnej funkcie	27
3.3.1 Odhad rozptylov	29
3.3.2 Závislé panely	29
4 Analýza reálnych dát	30
4.1 Dátový súbor	30
4.2 Popis modelov	30
4.3 Aplikácia WBS metódy na dáta	32
4.4 Reziduálna diagnostika	39
Záver	42
Zoznam použitej literatúry	43

Zoznam obrázkov	45
Zoznam tabuliek	46
Zoznam použitých skratiek	47

Úvod

Problematika detekcie bodov zmeny v časových radoch je štatistikmi skúmaná už niekoľko desaťročí. Pôvodne sa detekcia bodov zmeny študovala v kontexte kontroly kvality. V dnešnej dobe jej využitie nájdeme v ekonómii (James a Matteson, 2014; Bai, 2010; Fryzlewicz, 2014; Baranowski a kol., 2019), poisťovníctve (Peštová a Pešta, 2017; Pešta a kol., 2020; Maciak a kol., 2020), klimatológii (Baranowski a kol., 2019), v epidemiológii, medicíne a mnohých ďalších oblastiach. Všeobecne sa najčastejšie analyzuje časový rad, ktorý je popísaný stochastickým modelom s neznámymi parametrami. Tieto parametre môžu s rastúcim časovým indexom podliehať náhlym zmenám a byť tak po častiach konštantnou funkciou času. V záujme štúdie je potom určiť koľko a v akých časových indexoch potenciálne zmeny nastali, prípadne štatisticky otestovať, či zmena skutočne nastala. V práci sa sústreďujeme na detekciu bodov zmeny v strednej hodnote jednorozmerných časových radov a na rozšírenie problematiky na panelové dáta. Práca je členená nasledovne.

V prvej kapitole zavádzame modely časového radu s jedným a viacerými bodmi zmeny v strednej hodnote. Uvádzame testy o neprítomnosti bodu zmeny založené na CUSUM štatistikách a pomeroch vierohodností. Nakoniec sa zameriavame na odhadovanie počtu bodov zmeny pomocou Schwarzovho kritéria.

V druhej kapitole sa sústreďujeme na vysvetlenie novej metódy (Fryzlewicz, 2014) odhadovania počtu a polôh bodov zmeny, ktorá má pôvod v binárnej segmentácii. Uvádzame predpoklady fungovania tejto metódy, ktorú Fryzlewicz (2014) nazval WBS, a predpoklady pre fungovanie binárnej segmentácie. Ďalej vysvetľujeme algoritmus fungovania binárnej segmentácie a algoritmus fungovania WBS metódy s použitím penalizačného parametra. Nakoniec uvádzame zosilnené Schwarzovo informačné kritérium ako penalizačné kritérium pre WBS metódu (Fryzlewicz, 2014).

V tretej kapitole rozširujeme model časového radu zavedený v Kapitole 1 na panelové dáta. Vysvetľujeme odhadovanie bodu zmeny pomocou metódy najmenších štvorcov tak, ako ju popísal Bai (2010). Ďalej definujeme odhad bodu zmeny bez okrajovej podmienky (Peštová a Pešta, 2017). V ďalšej časti definujeme odhad bez okrajovej podmienky pre závislé panely (Pešta a kol., 2020; Maciak a kol., 2020). Uvedieme predpoklady pre konzistenciu zavedených odhadov, testové štatistiky založené na neparametrickom bootstrape, predpoklady pre ich konzistenciu a ich asymptotické rozdelenie. Nakoniec definujeme odhad bodu zmeny pomocou kvazi-vierohodnostnej funkcie, ktorý zaviedli Horváth a Hušková (2012).

Štvrtá kapitola je zameraná na analýzu reálnych dát. Metódu WBS, ktorej fungovanie sme vysvetlili v Kapitole 2, aplikujeme na rad týždenných logaritmických výnosov akciového indexu Dow Jones. Na tomto dátovom súbore porovnáваме rôzne voľby penalizačných kritérií WBS metódy a ich vplyv na počet a polohy odhadnutých bodov zmeny. Okrem pôvodného radu WBS metódu použijeme aj na rezíduá z GARCH modelu pre skúmané výnosy akciového indexu Dow Jones.

1. Štruktúrálné zmeny v časových radoch

1.1 Model časového radu s bodom zmeny

Nech \mathbb{Z} je množina celých čísel. Uvažujme T pozorovaní Y_1, \dots, Y_T z náhodného procesu $\{Y_t, t \in \mathbb{Z}\}$ v diskretnom čase s reálnymi hodnotami, ktorý sa dá zapísať ako

$$Y_t = \mu + \delta I[t > \kappa] + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (1.1)$$

kde μ je nemenná a nenáhodná úroveň radu, $\{\varepsilon_t, t \in \mathbb{Z}\}$ je reziduálna zložka, pre ktorú platí $E(\varepsilon_t) = 0$ a $E(\varepsilon_t^2) = \sigma^2$, a $I[\cdot]$ je indikátor. Model (1.1) je model časového radu s možnou prítomnosťou bodu zmeny. Parameter $\kappa \in \{1, \dots, T\}$ udáva čas, kedy nastala možná štruktúrálna zmena v časovom rade. Parameter δ meria veľkosť zmeny v strednej hodnote, ktorá nastala v čase κ . Variančný parameter $\sigma^2 \in (0, \infty)$ a parametre κ , δ a μ sú neznáme.

Definujme si teraz $\kappa_0 = 0$ a $\kappa_{m+1} = T$ pre $m > 1$. Pre prípad m štruktúrálnych zmien uvažujeme model

$$Y_t = \mu + \delta_1 I[t > \kappa_1] + \delta_2 I[t > \kappa_2] + \dots + \delta_m I[t > \kappa_m] + \varepsilon_t, \quad t = 1, \dots, T, \quad (1.2)$$

kde $0 = \kappa_0 < \kappa_1 < \kappa_2 < \dots < \kappa_m < \kappa_{m+1} = T$ sú body zmeny a δ_j je veľkosť zmeny v strednej hodnote medzi $(j+1)$ -vým a j -tým úsekom radu Y_t pre $j = 1, \dots, m$. Ak teda označíme μ_j strednú hodnotu radu Y_t v j -tom úseku, potom $\delta_j = \mu_{j+1} - \mu_j$ pre $j = 1, \dots, m$. Reziduálna zložka $\{\varepsilon_t, t \in \mathbb{Z}\}$ a parameter μ sú ako v modeli (1.1). Zrejme pre počet bodov zmeny platí $m < T$ a parameter m je neznámy. Model (1.2) je model s po častiach konštantnými strednými hodnotami. Počet pozorovaní T v modeloch (1.1) a (1.2) sa zvyčajne uvažuje pevný, alebo $T \rightarrow \infty$.

V nasledovných sekciách tejto kapitoly vychádzame z článku, ktorého autormi sú Aue a Horváth (2013).

1.2 CUSUM procedúry

Naším zámerom je testovať štruktúrálnu stabilitu modelu (1.1), prípadne modelu (1.2). Testovaním štruktúrálnych zmien sa v literatúre najčastejšie rozumie jeden z nasledovných bodov

1. Testovanie nulovej hypotézy

$$H_0 : \delta_j = 0, \quad j = 1, \dots, m, \quad (1.3)$$

teda test na nemennosť nepodmienenej strednej hodnoty Y_t v čase.

2. Test na nemennosť podmienenej strednej hodnoty Y_t . V prípade, že je známa informácia o štruktúre strednej hodnoty v podobe r -rozmerných vysvetľujúcich premenných $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ a $\mu = \boldsymbol{\beta}^T \mathbf{X}_t$, $\delta_j = \boldsymbol{\beta}_j^{*T} \mathbf{X}_t$, ide o test na nulovosť koeficientov $\boldsymbol{\beta}_j^*$

$$H_0 : \boldsymbol{\beta}_j^* = 0, \quad j = 1, \dots, m. \quad (1.4)$$

3. Testovanie štrukturálnej stability podmieneného rozptylu.

Hypotézy testujeme oproti alternatíve, že nulová hypotéza neplatí, teda že existuje bod $j \in \{1, \dots, m\}$, v ktorom nastala štrukturálna zmena v strednej hodnote, prípadne rozptyle. V tejto práci sa zameriavame na testovanie hypotézy (1.3).

1.2.1 Štrukturálne zmeny v nepodmienenej strednej hodnote

Chceme testovať nulovú hypotézu (1.3) konštantnej nepodmienenej strednej hodnoty oproti alternatíve, že aspoň raz v sledovanom období sa stredná hodnota zmenila. Aue a Horváth (2013) využíva testovú štatistiku založenú na preškálovaných CUSUM procesoch. Testy stability v zmysle testovania hypotézy (1.4) založené na CUSUM štatistikách (z angl. *cumulative sums*) popisuje vo svojej knihe Cipra (2008).

Preškálovaný CUSUM proces $Z_T = \{Z_T(x), x \in [0,1]\}$, ktorý pri konštrukcii testovej štatistiky využíva Aue a Horváth (2013) má tvar

$$Z_T(x) = \frac{1}{\sqrt{T}} \left(\sum_{t=1}^{\lfloor Tx \rfloor} Y_t - \frac{\lfloor Tx \rfloor}{T} \sum_{t=1}^T Y_t \right), \quad x \in [0, 1], \quad (1.5)$$

kde $\lfloor \cdot \rfloor$ je dolná celá časť čísla. Za platnosti hypotézy nulovosti koeficientov $\delta_j, j = 1, \dots, m$ dostávame

$$Z_T(x) = \frac{1}{\sqrt{T}} \left(\sum_{t=1}^{\lfloor Tx \rfloor} \varepsilon_t - \frac{\lfloor Tx \rfloor}{T} \sum_{t=1}^T \varepsilon_t \right), \quad x \in [0, 1],$$

teda CUSUM štatistika nezávisí na neznámom parametri μ .

Po dosadení bodu $x = k/T$ do CUSUM procesu (1.5) pre $k \in \{1, \dots, T\}$ dostávame štatistiku, ktorá porovnáva výberový priemer pozorovaní do času k , v ktorom nastala možná štrukturálna zmena, s výberovým priemerom pozorovaní cez celé sledované obdobie

$$Z_T(k/T) = \frac{k}{\sqrt{T}} \left(\frac{1}{k} \sum_{t=1}^k Y_t - \frac{1}{T} \sum_{t=1}^T Y_t \right). \quad (1.6)$$

Keďže skutočný čas κ , kedy zmena nastala, nepoznáme, je potrebné vypočítať hodnoty $Z_T(k/T)$ pre $\forall k \in \{1, \dots, T\}$. Ďalej hľadáme také k , pre ktoré je rozdiel vo výberových priemeroch (1.6) maximalizovaný. Aue a Horváth (2013) pre testovanie hypotézy (1.3) používajú maximovú testovú štatistiku preškálovanú odhadom parametru ω . Zhruba povedané, parameter ω udáva limitné chovanie

rozptylu reziduálneho výberového priemeru (tzv. *long-run variance*). Parameter ω je teda rovný $\omega = \lim_{T \rightarrow \infty} \text{var} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t$. Maximová testová štatistika má tvar

$$M_T = \frac{1}{\hat{\omega}_T} \max_{1 \leq k \leq T} \left| Z_T \left(\frac{k}{T} \right) \right|,$$

kde $\hat{\omega}_T$ je odhad parametru ω získaný z dát. Odhadovaním parametra ω sa zaoberáme v Kapitole 3.

Hypotézu (1.3) zamietame pre veľké hodnoty M_T . Bod k , ktorý maximalizuje hodnotu M_T je navyše kandidát na bod zmeny. Navyše poznáme limitné rozdelenie testovej štatistiky M_T , ktorého kvantilové hodnoty pre rôzne hladiny významnosti α sa dajú nájsť v tabuľkách. Dá sa ukázať, že

$$M_T \xrightarrow[T \rightarrow \infty]{\mathcal{D}} M = \sup_{0 \leq x \leq 1} |B(x)|,$$

kde $B(x) = W(x) - xW(1)$ a $\{W(x), x \in [0,1]\}$ je Wienerov proces v spojitom čase.

1.3 Testy pomerom vierohodností

Predpokladajme teraz, že máme náhodný proces $\{Y_t, t \in \mathbb{Z}\}$ a že rozdelenie Y_t závisí na neznámom parametri $\theta_t \in \mathbb{R}^p$. Nech boli napozorované náhodné veličiny Y_1, \dots, Y_T z tohto rozdelenia. V tomto prípade štruktúrna stabilita procesu zodpovedá nulovej hypotéze nemennosti parametru θ_t , teda

$$H_0 : \theta_1 = \dots = \theta_T = \theta_0. \quad (1.7)$$

Alternatíva potom vypadá nasledovne

$$H_1 : \theta_1 = \dots = \theta_\kappa = \theta_0 \neq \theta_0^* = \theta_{\kappa+1} = \dots = \theta_T,$$

teda v bode κ sa rozdelenie napozorovaných vektorov zmenilo. Skutočný čas κ , kedy táto zmena nastala, nepoznáme.

Jednou z možností testovania hypotézy (1.7) je test pomerom vierohodností. Za predpokladu, že zmena nastala v bode $k \in \{1, \dots, T\}$, sa dáta Y_1, \dots, Y_T rozdelia na dve podmnožiny, Y_1, \dots, Y_k s rozdelením závislým na θ_0 a Y_{k+1}, \dots, Y_T s rozdelením závislým na θ_0^* . Prípade $k = T$ zodpovedá situácii, keď zmena nastala. V testovej štatistike sa potom používa pomer vierohodností $L_k(\theta, \theta^*)$, ktorá berie v úvahu zmenu rozdelenia v bode k , a $L_T(\theta)$ pochádzajúcej z modelu za platnosti hypotézy, teda

$$\Lambda_k = \frac{\max_{\theta} L_T(\theta)}{\max_{\theta} L_k(\theta, \theta^*)} = \frac{L_T(\hat{\theta}_T)}{L_k(\hat{\theta}_k, \hat{\theta}_k^*)},$$

kde $\hat{\theta}_T$ je maximálne vierohodný odhad parametru θ_0 za platnosti nulovej hypotézy a $\hat{\theta}_k$ a $\hat{\theta}_k^*$ sú poporadí maximálne vierohodné odhady parametrov θ_0 a θ_0^* vo všeobecnom modeli s možnou zmenou v bode k .

Za testovú štatistiku pre test hypotézy (1.7) sa berie najvyššia hodnota negatívneho pomeru log-vierohodností cez všetky možné body zmeny k

$$Z_T = \max_{1 \leq k \leq T} (-2 \log \Lambda_k) = \max_{1 \leq k \leq T} 2(l_k(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\theta}}_k^*) - l_T(\hat{\boldsymbol{\theta}}_T)),$$

kde $l_k(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ a $l_T(\boldsymbol{\theta})$ sú log-vierohodnosti príslušných modelov. Hypotézu (1.7) zamietame pre veľké hodnoty testovej štatistiky Z_T . Pre $k = T$ je hodnota Z_T rovná nule.

1.4 Odhad počtu bodov zmeny

Počet bodov zmeny m v modeli (1.2) odhadujú Aue a Horváth (2013) pre dané delenie s potenciálnymi bodmi zmeny $0 = k_0 < k_1 < k_2 < \dots < k_h < k_{h+1} = T$ a pre $h \leq H$ použitím reziduálneho súčtu štvorcov

$$S(k_1, \dots, k_h) = \sum_{l=1}^{h+1} \sum_{t=k_{l-1}+1}^{k_l} (Y_t - \bar{Y}(k_{l-1}, k_l))^2,$$

kde $\bar{Y}(k_{l-1}, k_l) = (Y_{k_{l-1}+1} + \dots + Y_{k_l}) / (k_l - k_{l-1})$ je výberový priemer v l -tom úseku pre $l = 1, \dots, h+1$. Ďalej definujeme

$$\hat{\sigma}_h^2 = \frac{1}{T} \min S(k_1, \dots, k_h)$$

minimalizované cez jednotlivých kandidátov na body zmeny $0 = k_0 < k_1 < k_2 < \dots < k_h < k_{h+1} = T$. Yao (1988) odhaduje počet bodov zmeny m pomocou Schwarzovho kritéria ako

$$\hat{m} = \arg \min_{0 \leq h \leq H} SC(m), \quad SC(h) = \frac{T}{2} \log \hat{\sigma}_h^2 + h \log T.$$

Prvý člen v Schwarzovom kritériu meria kvalitu modelu pomocou log-vierohodnostnej funkcie a druhý člen slúži na penalizáciu odhadnutého počtu bodov zmeny. Dá sa ukázať, že za istých predpokladov je odhad \hat{m} slabozistentným odhadom skutočného počtu bodov zmeny m , teda

$$\hat{m} \xrightarrow[T \rightarrow \infty]{P} m.$$

2. Binárna segmentácia a WBS metóda

Jednou z metód odhadovania počtu bodov zmeny m a ich polôh $\kappa_1, \dots, \kappa_m$ v modeli (1.2) je takzvaná WBS metóda (z angl. *wild binary segmentation*). Túto pomerne novú techniku odhadovania navrhol Fryzlewicz (2014). Vychádza z už dlho používanej binárnej segmentácie, v ktorej sa, stručne povedané, dáta delia na dva a následne viac segmentov podľa odhadnutej lokácie bodu zmeny, ktorý je zároveň deliacim bodom daného dátového segmentu. Procedúra sa rekurzívne opakuje na vzniknutých segmentoch, ktoré sa následne môžu ďalej deliť v závislosti na prítomnosti ďalších bodov zmeny. Pri odhadovaní lokácií sa zvyčajne používa CUSUM štatistika.

Štandardná binárna segmentácia, aj napriek svojim mnoho výhodám, akými je napríklad nízka výpočetná náročnosť a celková jednoduchosť prevedenia, môže byť, ako ukazuje Fryzlewicz (2014), nevhodná na použitie pre niektoré modely s určitým rozložením bodov zmeny. WBS metóda má tieto nedostatky odstrániť, pričom si má zachovať nízku výpočetnú náročnosť a jednoduchosť prevedenia.

Hlavným rozdielom medzi hore uvedenými metódami je, že binárna segmentácia využíva v prvom kroku algoritmu celý dátový súbor (Y_1, \dots, Y_T) , ktorý potom rozdelí na dve časti podľa odhadnutého bodu zmeny, ktorý maximalizuje príslušnú CUSUM štatistiku. Vo WBS metóde sa náhodne vyberie určitý počet vektorov $(Y_s, Y_{s+1}, \dots, Y_e)$ pre $1 \leq s \leq e \leq T$, ktoré sú podmnožinami celého dátového súboru. Na týchto podmnožinách sa následne spočítajú príslušné CUSUM štatistiky, ktoré sú maximalizované cez všetky možné body zmeny k na danej podmnožine dát. Zo všetkých napočítaných kandidátov na bod zmeny sa potom vyberie ten s najvyššou hodnotou príslušnej CUSUM štatistiky. Podľa tohto bodu sa dáta rozdelia, podobne ako v binárnej segmentácii, na dve časti a ďalej sa postupuje rekurzívne.

Cieľom je dospieť k takému deleniu dátového súboru, ktoré umožňuje jednoducho a presne detekovať jednotlivé body zmeny, a to z relatívne nízkeho počtu náhodných výberov. Takýto interval (s, e) môže byť napríklad interval s práve jedným bodom zmeny, ktorý je dostatočne vzdialený od koncových bodov s a e (Fryzlewicz, 2014).

2.1 Model a predpoklady

Vychádzame z modelu (1.2), ktorý si pre väčšiu prehľadnosť zápisu v tomto kontexte prepíšeme do nasledovného tvaru

$$Y_t = f_t + \varepsilon_t, \quad t = 1, \dots, T, \quad (2.1)$$

kde f_t je po častiach konštantná stredná hodnota radu s m bodmi zmeny, ktorých polohy $0 = \kappa_0 < \kappa_1 < \dots < \kappa_m < \kappa_{m+1} = T$ a počet m sú neznáme.

Fryzlewicz (2014) ďalej zavádza nasledovný predpoklad.

- Predpoklad 1.** i. Náhodná postupnosť $\{\varepsilon_t\}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$;
ii. Postupnosť $\{f_t\}_t$ je ohraničená, teda $|f_t| < \bar{f} < \infty$ pre $t = 1, \dots, T$.

Prvý bod Predpokladu 1 by mohol byť rozšírený na závislé chybové zložky, či chybové zložky s rôznym rozdelením. V literatúre sa často pracuje napríklad s autoregresnými procesmi nižších rádov. Takisto by sa dal vypustiť predpoklad normality. Pre jednoduchosť sa však uvažujú i.i.d.chybové zložky s normálnym rozdelením. Predpoklad na jednotkový rozptyl nie je obmedzujúci, nakoľko $\text{var}(\varepsilon_t)$ je možné pomerne presne odhadnúť z dát za predpokladu, že máme dostatočný počet pozorovaní. Asi najtriviálnejším konzistentným odhadom rozptylu je výberový rozptyl. Fryzlewicz (2014) na odhadovanie rozptylu používa mediánovú absolútnu odchýlku.

Ďalej sa zavádzajú predpoklady na vzdialenosti medzi bodmi zmeny a na veľkosti skokov v stredných hodnotách. Predpoklady sa líšia pre štandardnú binárnu segmentáciu a pre metódu WBS.

Predpoklad 2. (Pre binárnu segmentáciu) Minimálne vzdialenosti medzi bodmi zmeny splňujú $\min_{1 \leq j \leq m+1} |\kappa_j - \kappa_{j-1}| \geq \eta_T$, kde $\eta_T \geq CT^\Theta$ pre $C > 0$ a $\Theta \leq 1$. Veľkosti skokov $\delta_j = |f_{\kappa_j} - f_{\kappa_{j-1}}|$ navyše splňujú $\min_{1 \leq j \leq m} \delta_j \geq \underline{f}_T$, kde $\underline{f}_T \geq CT^{-\bar{\omega}}$ pre $\bar{\omega} \geq 0$. Parametre Θ a $\bar{\omega}$ splňujú $\Theta - \frac{\bar{\omega}}{2} > \frac{3}{4}$.

Predpoklad 3. (Pre WBS) Minimálne vzdialenosti medzi bodmi zmeny splňujú $\min_{1 \leq j \leq m+1} |\kappa_j - \kappa_{j-1}| \geq \eta_T$ a veľkosti skokov $\delta_j = |f_{\kappa_j} - f_{\kappa_{j-1}}|$ splňujú $\min_{1 \leq j \leq N} \delta_j \geq \underline{f}_T$, kde η_T a \underline{f}_T splňujú $\eta_T^{1/2} \underline{f}_T \geq C \log^{1/2} T$ pre C dostatočne veľké.

Ďalšie podmienky na počet bodov zmeny m sa nezavádzajú a sú dané minimálnou vzdialenosťou medzi bodmi zmeny η_T . Špeciálne, ak to podmienka na vzdialenosti dovoľí, m môže stúpať do nekonečna s počtom pozorovaní T . Formálne môžeme písať, že m aj body zmeny κ_j sú funkciami T , teda $m = m(T)$ a $\kappa_j = \kappa_j(T)$ pre $j = 1, \dots, m + 1$.

2.2 Štandardná binárna segmentácia

Binárna segmentácia, podobne ako aj metóda WBS, vo svojom algoritme pri hľadaní kandidátov na body zmeny využíva CUSUM štatistiku. Tá je definovaná ako skalárny súčin vektoru (Y_s, \dots, Y_e) a príslušného vektoru váh

$$Z_{s,e}^k = \frac{1}{\sqrt{n}} \left(\sqrt{\frac{e-k}{(k-s+1)}} \sum_{t=s}^k Y_t - \sqrt{\frac{k-s+1}{(e-k)}} \sum_{t=k+1}^e Y_t \right), \quad (2.2)$$

kde $s \leq k < e$ a $n = e - s + 1$ je dĺžka úseku (Y_s, \dots, Y_e) . Kandidáti na body zmeny v jednotlivých iteráciách sa hľadajú maximalizovaním absolútnej hodnoty CUSUM štatistiky (2.2) cez jednotlivé body v danom intervale.

Ak hodnota CUSUM štatistiky (2.2) v nájdenom kandidátovi prekročí v absolútnej hodnote istú hranicu, danú parametrom ζ_T , zaradí sa tento kandidát medzi

odhadnuté body zmeny. Ten potom slúži ako deliaci bod v danom segmente a postup sa následne prevádza rekurzívne na vzniknutých segmentoch.

Predpokladajme, že parameter ζ_T je daný. Binárnu segmentáciu (ďalej aj „BS“) môžeme popísať pomocou nasledovného algoritmu.

Algoritmus fungovania BS

0. Polož $s := 1$ a $e := T$.
1. Ak $e - s < 1$ potom **skonči**. Inak prirad $k^0 := \arg \max_{k \in \{s, \dots, e-1\}} |Z_{s,e}^k|$ a pokračuj krokom 2.
2. Ak $|Z_{s,e}^{k^0}| > \zeta_T$ potom
 - pridaj k^0 do množiny odhadnutých bodov zmeny.
 - pokračuj krokom 1. pre $s = s$ a $e = k^0$.
 - pokračuj krokom 1. pre $s = k^0 + 1$ a $e = e$.

Inak **skonči**.

Nasledovná veta ukazuje, že odhady bodov zmeny získané binárnou segmentáciou sú za Predpokladov 1, 2 a určitých ďalších podmienok konzistentné.

Veta 1. *Nech Y_t je z modelu (2.1) a nech platia Predpoklady 1 a 2. Nech \hat{m} a $\hat{\kappa}_1 < \dots < \hat{\kappa}_{\hat{m}}$ sú poporadí odhadnutý počet bodov zmeny a ich odhadnuté polohy získané štandardnou binárnou segmentáciou. Nech parameter ζ_T spĺňa $\zeta_T = c_1 T^\theta$ kde $\theta \in (1 - \Theta, \Theta - 1/2 - \bar{\omega})$ ak $\Theta \in (\frac{3}{4}, 1)$, alebo $\zeta_T \geq c_2 \log^p T$ ($p > 1/2$) a $\zeta_T \leq c_3 T^\theta$ ($\theta < 1/2 - \bar{\omega}$) ak $\Theta = 1$, pre ľubovoľné kladné konštanty c_1, c_2, c_3 . Potom existujú kladné konštanty C, C_1 tak, že $P(\mathcal{A}_T) \geq 1 - C_1 T^{-1}$, kde*

$$\mathcal{A}_T = \left\{ \hat{m} = m; \max_{1 \leq j \leq m} |\hat{\kappa}_j - \kappa_j| \leq C \epsilon_T \right\}$$

pre $\epsilon_T = T^2 \eta_T^{-2} (\underline{f}_T)^{-2} \log T$.

2.3 WBS metóda

V tejto sekcii si vysvetlíme princíp fungovania WBS metódy, ktorú navrhol Fryzlewicz (2014). Výhodou WBS metódy oproti štandardnej binárnej segmentácii je, že má miernejšie podmienky na veľkosti skokov v stredných hodnotách a tiež na minimálne vzdialenosti medzi bodmi zmeny.

Nech F_T^G je množina G náhodných intervalov $[s_g, e_g], g = 1, \dots, G$, ktorých počiatkové a koncové body boli získané náhodným výberom s vracaním z diskrétného rovnomerného rozdelenia na množine $\{1, \dots, T\}$. Použitie náhodných intervalov miesto pevných je motivované faktom, že pri niektorých lokáciách bodov zmeny je okolo týchto bodov nutné zvoliť dostatočne úzky interval $[s, e]$ aby ich bolo možné rozpoznať. Pri použití náhodných intervalov je pravdepodobnosť, že sa takto úzky interval nachádza v množine F_T^G vždy kladná a môže byť dostatočne veľká (Fryzlewicz, 2014). Ukazuje sa však, že pre veľký počet intervalov G

je rozdiel v efektívite náhodných a deterministických dizajnov minimálny. Počet náhodných intervalov G by sa mal voliť čo najväčší v závislosti na výpočetnej kapacite. Fryzlewicz (2014) odporúča voliť $G = 5000$ pre dátové súbory nie väčšie ako niekoľko tisíc pozorovaní.

Fryzlewicz (2014) spomína, že jednou z výhod WBS metódy, podobne ako binárnej segmentácie, je jednoduchosť jej naprogramovania. Jej fungovanie popisuje nasledovný algoritmus.

Algoritmus fungovania WBS metódy

0. Polož $s := 1$ a $e := T$.

1. Ak $e - s < 1$ potom **skonči**. Inak

- označ $\mathcal{G}_{s,e}$ množinu takých indexov g , pre ktoré je interval $[s_g, e_g] \in F_T^G$ a $[s_g, e_g] \subseteq [s, e]$;
- ($\mathcal{G}_{s,e} := \mathcal{G}_{s,e} \cup \{0\}$, kde $[s_0, e_0] = [s, e]$ – tento krok je voliteľný);
- priradiť $(g^0, k^0) := \arg \max_{g \in \mathcal{G}_{s,e}, k \in \{s_g, \dots, e_g - 1\}} |Z_{s_g, e_g}^k|$ a pokračuj krokom 2.

2. Ak $|Z_{s_{g^0}, e_{g^0}}^{k^0}| > \zeta_T$ potom

- pridaj k^0 do množiny odhadnutých bodov zmeny;
- pokračuj krokom 1. pre $s = s$ a $e = k^0$;
- pokračuj krokom 1. pre $s = k^0 + 1$ a $e = e$;

Inak **skonči**.

Podmienka $[s_g, e_g] \subseteq [s, e]$ v prvom bode kroku 1 spolu so sekvenčnou povahou algoritmu zabraňujú tomu, aby boli niektoré body zmeny odhadnuté algoritmom viackrát a v iných polohách. Druhý bod prvého kroku algoritmu je voliteľný a zavádza sa, aby algoritmus preskúmal okrem náhodne vybraných podintervalov tiež celý interval $[s, e]$. Tento krok je užitočný v prípade, že $[s, e]$ obsahuje iba jeden bod zmeny. Vtedy sa ako optimálne ukazuje preskúmať celý interval.

Na rozdiel od binárnej segmentácie sú výstupom WBS algoritmu body zmeny v poradí, ktoré zodpovedá klesajúcej hodnote maxim $|Z_{s_g, e_g}^k|$. Takýto výstup vzniká v dôsledku maximálizácie CUSUM štatistiky cez $g \in \mathcal{G}_{s,e}$.

Fryzlewicz (2014) uvádza pre WBS algoritmus nasledovnú vetu o konzistencii získaných odhadov.

Veta 2. *Nech Y_t je z modelu (2.1) a nech platia Predpoklady 1 a 3. Nech \hat{m} a $\hat{\kappa}_1 < \dots < \hat{\kappa}_{\hat{m}}$ sú poporadí odhadnutý počet bodov zmeny a ich odhadnuté polohy získané WBS metódou. Potom existujú konštanty C, \bar{C} tak, že ak platí $C \log^{1/2} T \leq \zeta_T \leq \bar{C} \eta_T^{1/2} \underline{f}_T$, potom $P(\mathcal{A}_T) \geq 1 - C_1 T^{-1} - T \eta_T^{-1} (1 - \eta_T^2 T^{-2} / 9)^G$*

$$\mathcal{A}_T = \left\{ \hat{m} = m; \max_{1 \leq j \leq m} |\hat{\kappa}_j - \kappa_j| \leq C \log T (\underline{f}_T)^{-2} \right\}$$

pre určité kladné konštanty C a C_1 .

Ukazuje sa, že prípustné rozmedzie parametra ζ_T je vo WBS metóde vždy väčšie než pri štandardnej binárnej segmentácii. Táto vlastnosť robí WBS metódu robustnejšiu voči nevhodnému nastaveniu hodnoty ζ_T . Fryzlewicz (2014) volí $\zeta_T = C\sqrt{2}\log^{1/2} T$ pre nejaké $C > 0$. Konštanta C sa môže zvoliť prostredníctvom simulačnej štúdie tak, aby takáto voľba fungovala pre široké spektrum signálov. Fryzlewicz (2014) berie za prednastavenú hodnotu $C = 1$ a ďalej potom $C = 1,3$. Rád $\zeta_T = O(\log^{1/2} T)$ je, ako tvrdí Fryzlewicz (2014), jediný, pomocou ktorého dostaneme konzistentné odhady pre ľubovoľnú prípustnú vzdialenosť η_T a minimálnu veľkosť skokov \underline{f}_T . Pre binárnu segmentáciu je jediný rád vedúci ku konzistencii odhadov rovný $T^{1/4-\bar{\omega}/2}$, kde $\bar{\omega}$ je z Vety 1.

Všimnime si, že množina \mathcal{A}_T vo Vete 2 nezávisí na minimálnej vzdialenosti η_T , ako je tomu vo Vete 1, ale iba na minimálnej veľkosti skokov \underline{f}_T . Splnená ale musí byť podmienka na η_T z Predpoladu 3.

Počet náhodných výberov G je nutné zvoliť tak, aby pravdepodobnosť $P(\mathcal{A}_T)$ vo Vete 2 bola dostatočne blízko jednej. Fryzlewicz (2014) uvažuje

$$T\eta_T^{-1} \left(1 - \eta_T^2 T^{-2}/9\right)^G \leq T^{-1}, \quad (2.3)$$

odkiaľ potom

$$\begin{aligned} P(\mathcal{A}_T) &\geq 1 - C_1 T^{-1} - T^{-1} \\ 1 - P(\mathcal{A}_T) &\leq T^{-1}(1 + C_1). \end{aligned}$$

Výraz (2.3) ďalej upravíme nasledovne

$$\begin{aligned} T^2\eta_T^{-1} &\leq \exp\left\{-G \log\left(1 - \eta_T^2 T^{-2}/9\right)\right\} \\ \log\left(T^2\eta_T^{-1}\right) &\leq -G \log\left(1 - \eta_T^2 T^{-2}/9\right), \end{aligned}$$

kde sa využije aproximácia $\log(1 - y) \approx -y$ na okolí $y = 0$, teda

$$\log\left(T^2\eta_T^{-1}\right) \leq G\eta_T^2 T^{-2}/9$$

a odtiaľ sa potom úpravami dostane dolná hranica pre G

$$G \geq \log\left(T^2\eta_T^{-1}\right) 9T^2\eta_T^{-2}.$$

Vidíme, že čím je minimálna vzdialenosť η_T menšia, tým vyšší počet náhodných výberov G je potrebných k získaniu konzistentných odhadov. Pomerne nízka výpočetná náročnosť sa dosiahne pre $\eta_T = O(T)$, kedy stačí logaritmický počet náhodných výberov.

2.3.1 Zosilnené Schwarzovo informačné kritérium pre WBS

Pri výbere množiny odhadnutých bodov zmeny zohráva veľkú úlohu voľba parametra ζ_T . Na veľkosti tohto parametra závisí počet odhadnutých bodov zmeny \hat{m} , ktorý je jeho neklesajúcou funkciou, teda $\hat{m} = \hat{m}(\zeta_T)$, a tiež samotné polohy

$\hat{\kappa}_1, \dots, \hat{\kappa}_m$. Označme $C(\zeta_T) = \{\hat{\kappa}_1, \dots, \hat{\kappa}_m\}$ množinu odhadnutých bodov zmeny. Predpokladajme, že pre nejaké H pevné také, že $m \leq H$ je $\{\zeta_T^h\}_{h=0}^H$ klesajúca postupnosť parametrov a $|C(\zeta_T^h)| = h$. Jednou z možností je pracovať s jedinou vhodne zvolenou hodnotou parametra ζ_T , pričom výstupom procedúry bude množina $C(\zeta_T)$. Alternatívne sa dá skúmať celá postupnosť $\{\zeta_T^h\}_{h=0}^H$ a príslušné množiny $\{C(\zeta_T^h)\}_{h=0}^H$. Najvhodnejšia množina $C(\zeta_T^h) = C_h$ sa volí optimalizovaním hodnoty nejakého kritéria. Fryzlewicz (2014) k tomuto účelu využíva takzvané zosilnené Schwarzovo informačné kritérium.

Pre ľubovoľný model C_h Fryzlewicz (2014) zavádza

$$\hat{f}_t^h = (\hat{\kappa}_{j+1} - \hat{\kappa}_j)^{-1} \sum_{l=\hat{\kappa}_j+1}^{\hat{\kappa}_{j+1}} Y_l, \quad \hat{\kappa}_j + 1 \leq t \leq \hat{\kappa}_{j+1}$$

ako odhad signálu f_t a

$$\hat{\sigma}_h^2 = T^{-1} \sum_{t=1}^T (Y_t - \hat{f}_t^h)^2$$

príslušný maximálne vierohodný odhad reziduálneho rozptylu. Fryzlewicz (2014) ďalej definuje zosilnené Schwarzovo informačné kritérium ako

$$sSIC(h) = \frac{T}{2} \log \hat{\sigma}_h^2 + h \log^\alpha T, \quad (2.4)$$

pričom $\alpha = 1$ zodpovedá štandardnému Schwarzovmu informačnému kritériu. V tomto kontexte sa požaduje $\alpha > 1$, čo vedie k silnejšej penalizácii odhadnutého počtu bodov zmeny než pri štandardnom Schwarzovom kritériu.

Veta 3. *Nech Y_t je z modelu (2.1) a nech platia predpoklady Vety 2. Nech m a $\kappa_1, \dots, \kappa_m$ sú poporadí počet a polohy bodov zmeny. Nech $m \leq H$, kde H je daná kladná konštanta nezávislá na m . Nech $\alpha > 1$ je taká konštanta, pre ktorú $\log^\alpha T = o(\eta_T \underline{f}_T^2)$. Nech modely $\{C_h\}_{h=0}^H$ boli získané WBS algoritmom a nech $\hat{m} = \arg \min_{0 \leq h \leq H} sSIC(h)$. Potom $P(\mathcal{A}_T) \geq 1 - C_1 T^{-1} - T \eta_T^{-1} (1 - \eta_T^2 T^{-2} / 9)^G$, kde*

$$\mathcal{A}_T = \left\{ \hat{m} = m; \max_{1 \leq j \leq m} |\hat{\kappa}_j - \kappa_j| \leq C \log T (\underline{f}_T)^{-2} \right\}$$

pre určité kladné konštanty C a C_1 .

Keďže sa vo Vete 3 požaduje, aby $\log^\alpha T = o(\eta_T \underline{f}_T^2)$, pracuje sa s α blízkym jednej, aby sa zaistila prípustnosť kritéria pre čo najväčšiu množinu signálov f_t . Odhad \hat{m} sa získa podobne ako v Sekcii 1.4, tentokrát minimalizáciou kritéria (2.4).

3. Odhady bodov zmeny pre panelové dáta

Model (1.1) s jedným skúmaným radom rozšírime pre prípad N radov (panelov), teda

$$Y_{i,t} = \mu_i + \delta_i I[t > \kappa] + \varepsilon_{i,t}, \quad 1 \leq i \leq N, 1 \leq t \leq T, \quad (3.1)$$

kde podobne ako v pôvodnom modeli (1.1) je μ_i stredná hodnota i -teho radu do času zmeny κ a $\{\varepsilon_{i,t}, 1 \leq t \leq T\}$ sú reziduálne zložky s nulovými strednými hodnotami a druhými momentmi rovnými $E(\varepsilon_{i,t}^2) = \sigma_i^2$ pre $\sigma_i \in (0, \infty)$ a $i = 1, \dots, N$ a $t = 1, \dots, T$. Bod zmeny $\kappa \in \{1, \dots, T\}$ je rovnaký pre všetky panely.

Veľkosť zmeny v strednej hodnote δ_i sa môže v jednotlivých paneloch líšiť. Parameter δ_i je deterministický, alebo pochádza z nejakého pravdepodobnostného rozdelenia. Model (3.1) je model s nanaajvyš jedným bodom zmeny na jeden panel, avšak podobne ako model (1.1) môže byť zovšeobecnený pre $m > 1$ bodov zmeny. Parametre $\kappa, \delta_i, \sigma_i$ a m sú podobne ako aj v modeli (1.2) neznáme.

Vo všeobecnom modeli predpokladáme, že panely sú nezávislé. Počet skúmaných panelov N býva zvyčajne väčší než je počet pozorovaní T . V prípade, že pracujeme s mikroekonomickými dátami, keď N reprezentuje počet firiem, či osôb a T reprezentuje čas, môže byť N rádovo vyššie než T .

Parameter T sa uvažuje pevný alebo $T \rightarrow \infty$. Pre pevné T môže parameter κ nadobúdať ľubovoľnú hodnotu z intervalu $[1, T-1]$. Niektorí autori odhadujú bod zmeny bez okrajovej podmienky, takže κ môže nadobúdať aj hodnotu T na konci sledovaného intervalu (Peštová a Pešta, 2017; Pešta a kol., 2020; Maciak a kol., 2020). Ak uvažujeme $T \rightarrow \infty$, potom sa pre bod zmeny predpokladá, že je tvaru $\kappa = \lfloor Tx \rfloor$ pre $x \in (0, 1)$. Bai (2010) ukázal, že tento predpoklad na κ sa dá odstrániť, ak platí, že N je oveľa väčšie než T , teda ak $T/N \rightarrow 0$ pre $N, T \rightarrow \infty$.

V tomto modeli sa testuje hypotéza

$$H_0 : \delta_i = 0 \quad \forall i = 1, \dots, N, \quad (3.2)$$

že zmena nenastala pre žiadny z N panelov oproti alternatíve

$$H_1 : \exists i^* \in \{1, \dots, N\} : \delta_{i^*} \neq 0,$$

že zmena nastala aspoň v jednom paneli. V nasledujúcej sekcii vychádzame z článku Bai (2010).

3.1 Odhad bodu zmeny metódou najmenších štvorcov

Bai (2010) popisuje odhadovanie bodu zmeny κ pomocou najmenších štvorcov pre panelové dáta za podmienky, že zmena určite nastala. Odhad $\hat{\kappa}_{LS}$ bodu zmeny

κ sa preto hľadá v intervale $[1, T - 1]$ a definuje sa ako

$$\hat{\kappa}_{LS} = \arg \min_{1 \leq k \leq T-1} \sum_{i=1}^N \left(\sum_{t=1}^k (Y_{i,t} - \bar{Y}_{i,k})^2 + \sum_{t=k+1}^T (Y_{i,t} - \tilde{Y}_{i,k})^2 \right), \quad (3.3)$$

kde $\bar{Y}_{i,k}$ je výberový priemer prvých k pozorovaní a $\tilde{Y}_{i,k}$ je výberový priemer posledných $T - k$ pozorovaní v i -tom paneli, teda

$$\bar{Y}_{i,k} = \frac{1}{k} \sum_{t=1}^k Y_{i,t} \quad \text{a} \quad \tilde{Y}_{i,k} = \frac{1}{T-k} \sum_{t=k+1}^T Y_{i,t}. \quad (3.4)$$

Bai (2010) navyše ukazuje, že odhad (3.3) je konzistentným odhadom skutočného bodu zmeny κ v zmysle slabšej konzistencie za platnosti nasledovných predpokladov.

- Predpoklad 4.**
- i. $\varepsilon_{i,t} = \sum_{j=0}^{\infty} a_{i,j} e_{i,t-j}$, kde $\{a_{i,j}, j \in \mathbb{N}\}$ je postupnosť konštant;
 - ii. Pre každé i sú $\{e_{i,t}\}_t \sim (0, \sigma_{i,e}^2)$ i.i.d.;
 - iii. $\sum_{j=0}^{\infty} j |a_{i,j}| \leq M$ pre $\forall i$ a $M < \infty$;
 - iv. Postupnosti $\{\varepsilon_{i,t}\}_t$ sú nezávislé cez i .

Podľa prvých troch bodov Predpokladu 4 tvoria reziduálne zložky $\varepsilon_{i,t}$ lineárny proces. Niekedy sa v definícii lineárneho procesu miesto tretieho bodu Predpokladu 4 uvádza podmienka $\sum_{j=0}^{\infty} |a_{i,j}| < \infty$ (Prášková, 2004). Štvrtý bod Predpokladu 4 hovorí o nezávislosti reziduálnych zložiek medzi panelmi. Tento bod je možné z predpokladov vypustiť, avšak Bai (2010) pracuje pre jednoduchosť s nezávislosťou. Predpokladá sa, že rozptyl reziduálnych zložiek panelov je nemenný a $\sigma_i^2 = \mathbb{E}(\varepsilon_{i,t})^2 = \sigma_{i,e}^2 \sum_{j=0}^{\infty} a_{i,j}^2$.

Ďalší predpoklad udáva podmienku na veľkosti zmien δ_i v stredných hodnotách.

Predpoklad 5.

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \sum_{i=1}^N \delta_i^2 = \infty.$$

Podľa Predpokladu 5 musia byť zmeny v stredných hodnotách dostatočne veľké, aby sme dokázali získať konzistentný odhad bodu zmeny κ .

Veta 4. *Nech pre model (3.1) platia Predpoklady 4 a 5. Potom pre T pevné alebo pre $T \rightarrow \infty$ platí*

$$\lim_{N \rightarrow \infty} P(\hat{\kappa}_{LS} = \kappa) = 1.$$

Podrobný dôkaz Vety 4 a ďalších viet v tejto a nasledujúcej sekcii sa nachádza v prílohe článku Bai (2010).

Bai (2010) navyše ukazuje, že za ďalšieho predpokladu na počet pozorovaní T je možné uvažovať slabšiu podmienku na veľkosti zmien δ_i než je podmienka v Predpoklade 5. Počet pozorovaní T musí byť dostatočne väčší než je počet panelov N a to tak, aby bol splnený nasledovný predpoklad.

Predpoklad 6.

$$\lim_{N, T \rightarrow \infty} \frac{\log(\log(T))}{T} N = 0.$$

Za takýchto podmienok sa potom uvažuje slabšia podmienka na divergenciu súčtu štvorcov veľkostí δ_i , pri ktorej už nepožadujeme rýchlosť, s akou by táto suma mala divergovať.

Predpoklad 7.

$$\lim_{N \rightarrow \infty} \sum_{i=1}^N \delta_i^2 = \infty.$$

Za Predpokladov 4, 6 a 7 je odhad $\hat{\kappa}_{LS}$ konzistentným odhadom bodu zmeny κ , a to pri dostatočnom počte pozorovaní T a panelov N .

Veta 5. *Za predpokladov 4, 6 a 7 dostávame*

$$\lim_{N, T \rightarrow \infty} P(\hat{\kappa}_{LS} = \kappa) = 1.$$

3.1.1 Limitné rozdelenie

V predchádzajúcej sekcii sme si ukázali podmienky, za ktorých odhad $\hat{\kappa}_{LS}$ konverguje v pravdepodobnosti ku skutočnému bodu zmeny κ . V praxi sú ale tieto podmienky často ťažko dosiahnuteľné. Dôvodom býva nedostatočné množstvo pozorovaných panelov, či príliš malé veľkosti zmien δ_i . Z týchto dôvodov sa skúma limitné rozdelenie odhadu $\hat{\kappa}_{LS}$, na základe ktorého je potom možné konštruovať intervaly spoľahlivosti pre κ .

Bai (2010) pri konštrukcii predpokladá, že T aj N divergujú do nekonečna, keďže pre T pevné závisí limitné rozdelenie na presnej polohe skutočného bodu zmeny κ a má komplikovanejšie vyjadrenie.

Namiesto Predpokladu 7 sa požaduje, aby súčet štvorcov magnítud konvergoval pri dostatočnom počte panelov N k nejakej konštantne λ .

Predpoklad 8.

$$\lim_{N \rightarrow \infty} \sum_{i=1}^N \delta_i^2 = \lambda, \quad \lambda \in (0, \infty).$$

V nasledovnom predpoklade si zhrnieme ďalšie podmienky na magnítudy δ_i , ktoré pri odvodzovaní limitného rozdelenia uvažuje Bai (2010).

Predpoklad 9. Pre Δ_i rovnomerne ohraničené (alebo v prípade stochastického Δ_i pre Δ_i s rovnomerne ohraničeným rozptylom) je

1. $\delta_i = N^{-1/2}\Delta_i$
2. $\lim_{N \rightarrow \infty} \sum_{i=1}^N \delta_i^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \Delta_i^2 = \lambda, \quad \lambda \in (0, \infty)$
3. $\lim_{N \rightarrow \infty} \sum_{i=1}^N \delta_i^2 \sigma_i^2 = \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \Delta_i^2 \sigma_i^2 = \phi.$

Nasledovná Lemma, ktorú využíva Bai (2010) k dôkazu limitného rozdelenia $\hat{\kappa}_{LS}$, hovorí, že rozdiel odhadu $\hat{\kappa}_{LS}$ a jeho skutočnej hodnoty je stochasticky ohraničený, teda platí, že pre všetky $\varepsilon > 0$ existuje konečné $K > 0$ a konečné $L > 0$ tak, že

$$\mathbb{P}(|\hat{\kappa}_{LS} - \kappa| > K) < \varepsilon, \forall n > L.$$

Lemma 6. *Za Predpokladov 4, 6 a bodov 1 a 2 Predpokladu 9 platí*

$$\hat{\kappa}_{LS} - \kappa = O_p(1).$$

Ak podelíme odhad $\hat{\kappa}_{LS}$ počtom pozorovaní T , z Lemmy 6 dostávame T -konzistentný odhad. Pre $T \rightarrow \infty$ a $\hat{\tau} = \hat{\kappa}_{LS}/T$ máme

$$\hat{\tau} = \tau + O_p(T^{-1}),$$

pre $\tau = \kappa/T$.

Výsledky o limitnom rozdelení $\hat{\kappa}_{LS}$, ktoré odvodil Bai (2010) sú zhrnuté v nasledujúcej Vete.

Veta 7. *Nech $\varepsilon_{i,t}$ sú nekorelované. Potom za Predpokladov 6 a 9 pre $N, T \rightarrow \infty$ máme*

$$\hat{\kappa}_{LS} - \kappa \xrightarrow{\mathcal{D}} \arg \min_l \left\{ |l| \lambda + 2\sqrt{\phi} W(l) \right\}, \quad (3.5)$$

kde $W(0) = 0$ a

$$W(l) = \sum_{s=-l+1}^0 Z_s, \quad l = -1, -2, \dots$$

$$W(l) = \sum_{s=1}^l Z_s, \quad l = 1, 2, \dots$$

a $Z_s, s = \dots - 2, -1, 0, 1, 2, \dots$ sú *i.i.d.* náhodné veličiny so štandardným normálnym rozdelením.

Veta 7 platí aj pre korelované reziduálne zložky $\varepsilon_{i,t}$ s tým rozdielom, že v takom prípade budú korelované aj náhodné veličiny Z_s . Vtedy je nutné odhadovať korelácie $\mathbf{E}(Z_s Z_{s+h})$. Veličina $W(l)$ vystupujúca v limitnom rozdelení $\hat{\kappa}_{LS}$ je náhodná prechádzka v diskretnom čase.

Pre model (1.1) z Kapitoly 1 s jedným časovým radom a za predpokladov, že pre $T \rightarrow \infty$ platí $\delta \rightarrow 0$ a $T\delta^2 \rightarrow \infty$, je limitné rozdelenie odhadu $\hat{\kappa}_{LS}$ získaného metódou najmenších štvorcov

$$(\delta^2/\omega^2)(\hat{\kappa}_{LS} - \kappa) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \arg \min_r \left\{ -|r| + 2W(r) \right\}, \quad (3.6)$$

kde ω je ako v Kapitole 1. $W(r)$ tu vystupuje ako Brownovský pohyb (tj. náhodná prechádzka v spojitom čase) na množine reálnych čísel, ktorý navyše splňuje $W(0) = 0$, $W(r) = W_1(r)$ pre $r > 0$ a $W(r) = W_2(-r)$ pre $r < 0$, kde $W_1(r)$ a $W_2(r)$ sú dva nezávislé Brownovské pohyby na $[0, \infty)$.

Ak by sme v limitnom rozdelení (3.5) náhodnú prechádzku $W(l)$ v diskretnom čase nahradili Brownovským pohybom, limitné rozdelenia (3.5) a (3.6) by sa líšili iba o parametre λ a ϕ vystupujúce v limitnom rozdelení (3.5) a konštantu -1 . Limitné rozdelenie potom môžeme prepísať nasledovne

$$(\lambda^2/\phi)(\hat{\kappa}_{LS} - \kappa) \xrightarrow{\mathcal{D}} \arg \min_l \{|l| + 2W(l)\} \quad (3.7)$$

pre $N, T \rightarrow \infty$. Bai (2010) ďalej značí

$$A_N = \frac{\left(\sum_{i=1}^N \delta_i^2\right)^2}{\sum_{i=1}^N \delta_i^2 \sigma_i^2}.$$

Vzťah (3.7) sa potom dá prepísať ako

$$A_N(\hat{\kappa}_{LS} - \kappa) \xrightarrow{\mathcal{D}} \arg \min_l \{|l| + 2W(l)\}, \quad (3.8)$$

keďže pre $N \rightarrow \infty$ je $A_N \rightarrow \lambda^2/\phi$ z bodov 2 a 3 Predpokladu 9.

Limita (3.8) by platila pre l spojité. V tomto kontexte sa ale uvažuje $l \in \mathbb{Z}/\{0\}$, preto je vzťah (3.8) iba aproximatívny. Náhodná prechádzka v diskretnom čase a Brownovský pohyb pre čas $t \in \mathbb{Z}$, na ktorom sa táto aproximácia zakladá, majú ale rovnaké rozdelenie, preto je veľmi presná. Navyše aproximatívne rozdelenie (3.8) nezávisí na neznámych parametroch.

Asymptotický interval spoľahlivosti s pokrytím $(1 - \alpha)$ založený na vzťahu (3.8) je

$$\left(\hat{\kappa}_{LS} - \lfloor l_{1-\alpha/2}^*/\hat{A}_N \rfloor, \hat{\kappa}_{LS} + \lceil l_{1-\alpha/2}^*/\hat{A}_N \rceil\right), \quad (3.9)$$

kde $l_{1-\alpha/2}^*$ je $(1 - \alpha/2)$ - kvantil rozdelenia náhodnej veličiny

$$l^* = \arg \min_l \{|l| + 2W(l)\}.$$

Bai (2010) nasimuloval kvantilové hodnoty $l_{1-\alpha/2}^*$ pre hladiny spoľahlivosti $\alpha \in \{0,1; 0,05, 0,01\}$. Odhad \hat{A}_N v intervale spoľahlivosti (3.9) je

$$\hat{A}_N = \frac{\left(\sum_{i=1}^N \hat{\delta}_i^2\right)^2}{\sum_{i=1}^N \hat{\delta}_i^2 \hat{\sigma}_i^2},$$

kde $\hat{\delta}_i$ a $\hat{\sigma}_i$ sa získajú z dát.

Bai (2010) ukazuje, že rozdiel $\hat{A}_N - A_N$ konverguje v pravdepodobnosti k nule pre $N \rightarrow \infty$, teda

$$\hat{A}_N - A_N = o_p(1),$$

kde rádom $o_p(1)$ rozumieme, že

$$\lim_{N \rightarrow \infty} \mathbf{P}(|\hat{A}_N - A_N| \geq \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

Teda interval spoľahlivosti (3.9) asymptoticky pokrýva skutočný bod zmeny κ s pravdepodobnosťou $1 - \alpha$.

3.2 Odhad bodu zmeny bez okrajovej podmienky

V tejto sekcii popíšeme metódu odhadovania bodu zmeny κ , ktorú navrhli Peštová a Pešta (2017) a vychádza z vyššie popísanej metódy najmenších štvorcov. Na rozdiel od metódy, s ktorou prišiel Bai (2010), sa nepredpokladá, že zmena vo výbere musela nastať. Získané odhady budú konzistentné bez ohľadu na prítomnosť bodu zmeny a to za pomerne slabých predpokladov, ako si ukážeme nižšie. Metóda je navyše schopná rozpoznať bod zmeny, ktorý sa nachádza blízko hraníc intervalu pozorovaní, a teda tu nie je prítomná žiadna okrajová podmienka na bod κ .

Vychádza sa z modelu (3.1) s nanaajvyš jedným bodom zmeny pre T pevné a nezávislé na počte panelov N . Prípád $\kappa = T$ tu znamená, že zmena v uvažovanom výbere nenastala.

Peštová a Pešta (2017) nepredpokladajú, že $\{\varepsilon_{i,t}\}_t$ tvoria lineárny proces, ako sa uvádza v Predpoklade 4. Postačujúce je, aby $\{\varepsilon_{i,t}\}_t$ boli i.i.d. a v každom paneli tvorili slabo stacionárnu postupnosť.

Predpoklad 10. Pre $\forall i = 1, \dots, N$ a pre $k, v = 1, \dots, T$ platí

1. vektory $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,T})$ sú i.i.d. a $\varepsilon_{i,k} \sim (0, 1)$;
2. autokorelačná funkcia $\rho_k = \text{Cor}(\varepsilon_{i,t}, \varepsilon_{i,t+k}) = \text{Cov}(\varepsilon_{i,t}, \varepsilon_{i,t+k})$ nezávisí na čase t , $\forall t = 1, \dots, T - k$;
3. kumulatívna autokorelačná funkcia vektorov $\boldsymbol{\varepsilon}_i$ je v tvare $r(k) = \text{var} \sum_{t=1}^k \varepsilon_{i,t} = \sum_{|t|<k} (k - |t|)\rho_t$;
4. posunutá kumulatívna autokorelačná funkcia vektorov $\boldsymbol{\varepsilon}_i$ je v tvare $R(k, v) = \text{Cov}(\sum_{t=1}^k \varepsilon_{i,t}, \sum_{u=k+1}^v \varepsilon_{i,u}) = \sum_{t=1}^k \sum_{u=k+1}^v \rho_{u-t}$, $k < v$;
5. variančná matica $\boldsymbol{\Lambda} := \text{var} [\varepsilon_{1,1}, \sum_{t=1}^2 \varepsilon_{1,t}, \dots, \sum_{t=1}^T \varepsilon_{1,t}]$ je regulárna.

Jednou z možných alternatív modelovania skúmaných časových radov k lineárnym procesom sú GARCH procesy. Predpoklad nezávislosti panelov by sa mohol znova vypustiť, pre jednoduchosť sa ale uvažuje nezávislosť.

Ďalším predpokladom, ktorý uvažujú Peštová a Pešta (2017), je predpoklad na rovnako ohraničené rozptyly σ_i jednotlivých panelov.

Predpoklad 11. Existujú konštanty $\underline{\sigma}, \bar{\sigma} > 0$, ktoré nezávisia na N , tak, že

$$\underline{\sigma} \leq \sigma_i \leq \bar{\sigma}, \quad 1 \leq i \leq N.$$

Vďaka Predpokladu 11 sa nevyžadujú homoskedastické panely, stačí, aby boli ich rozptyly zhora a zdola ohraničené rovnakými konštantami. Tento predpoklad umožňuje aplikovať navrhovanú metódu na omnoho širšie spektrum dát.

Odhad bodu zmeny κ , ktorý navrhli Peštová a Pešta (2017) je definovaný ako

$$\hat{\kappa}_P := \arg \min_{1 \leq k \leq T} \sum_{i=1}^N \left\{ \frac{1}{w(k)} \sum_{t=1}^k (Y_{i,t} - \bar{Y}_{i,k})^2 + \frac{1}{w(T-k)} \sum_{t=k+1}^T (Y_{i,t} - \tilde{Y}_{i,k})^2 \right\},$$

kde pre $k = 1, \dots, T-1$ je $\bar{Y}_{i,k}$ výberový priemer prvých k pozorovaní a $\tilde{Y}_{i,k}$ je výberový priemer posledných $T-k$ pozorovaní, oba definované v (3.4), a $\{w(k)\}_{k=0}^T$ je postupnosť kladných váh. Pre $k = T$ položíme

$$\bar{Y}_{i,k} = \frac{1}{T} \sum_{t=1}^T Y_{i,t}, \quad \tilde{Y}_{i,k} = 0. \quad (3.10)$$

Ako sme uviedli vyššie, Peštová a Pešta (2017) nepredpokladajú, že zmena nutne nastala. Ak teda zmena vo výbere nenastala, odhad $\hat{\kappa}_P$ bude rovný číslu T s pravdepodobnosťou bližiacou sa jednej.

Ak chceme dostať konzistentný odhad $\hat{\kappa}_P$, váhy $\{w(k)\}_{k=0}^T$ musia byť zvolené tak, aby boli splnené nasledovné predpoklady. Navyše sa požaduje konečnosť štvrtých momentov rezíduí.

Predpoklad 12. 1. Nech $g(k) = \frac{1}{w(k)} \left(1 - \frac{r(k)}{k^2}\right)$ pre $k \in \{1, \dots, T\}$, $g(0) \equiv 0$ a nech

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \left\{ \frac{\kappa}{\kappa+1} \sum_{i=1}^N \delta_i^2 - (g(\kappa) + g(T-\kappa)) \max_{1 \leq k \leq T} w(k) \sum_{i=1}^N \sigma_i^2 \right\} &= \infty \\ \lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \left\{ \frac{T-\kappa}{T-\kappa+1} \sum_{i=1}^N \delta_i^2 - (g(\kappa) + g(T-\kappa)) \max_{1 \leq k \leq T} w(k) \sum_{i=1}^N \sigma_i^2 \right\} &= \infty; \end{aligned}$$

$$2. \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N \delta_i^2 = 0;$$

$$3. E \varepsilon_{1,k}^4 < \infty, \quad k \in \{1, \dots, T\}.$$

Prvé dva body tohto predpokladu mimo iného kontrolujú veľkosti magnítúd δ_i , ktoré musia byť dostatočne veľké na to, aby boli detekovateľné, no zároveň nie priveľké, aby bol splnený bod 2. Za týchto podmienok Peštová a Pešta (2017) odvodili konzistenciu odhadu $\hat{\kappa}_P$ pre $N \rightarrow \infty$ a T pevné.

Veta 8. Za Predpokladov 10 – 12,

$$\lim_{N \rightarrow \infty} P(\hat{\kappa}_P = \kappa) = 1.$$

Prvý bod Predpokladu 12 pojednáva o vzťahu veľkosti magnítúd δ_i a rozptylov σ_i^2 chybových zložiek. Pri väčšej variabilite sa vyžadujú väčšie magnítúdy δ_i a naopak. Táto podmienka tiež hovorí o tom, ako správne zvoliť váhy $w(k)$, aby bol získaný odhad konzistentný, napríklad aj v závislosti na hodnote autokorelácií $r(k)$.

Nulová hypotéza (3.2) o neprítomnosti bodu zmeny v jednotlivých paneloch sa bez prítomnosti okrajovej podmienky na κ dá písať aj nasledovne

$$H_0 : \kappa = T.$$

oproti alternatíve

$$H_1 : \kappa < T \quad \text{a} \quad \exists i^* \in \{1, \dots, N\} : \delta_{i^*} \neq 0.$$

Jednou z možností testovania takejto nulovej hypotézy je použitie podielovej maximovej testovej štatistiky tvaru

$$S_{N,T} = \max_{k=2,\dots,T-2} \frac{\max_{t=1,\dots,k-1} |L_N(t, k)|}{\max_{t=k,\dots,T-1} |R_N(t, k)|}, \quad (3.11)$$

kde $L_N(t, k) := \sum_{i=1}^N \sum_{r=1}^t (Y_{i,r} - \bar{Y}_{i,k})$, $R_N(t, k) = \sum_{i=1}^N \sum_{r=t+1}^T (Y_{i,r} - \tilde{Y}_{i,k})$ a $\bar{Y}_{i,k}$, $\tilde{Y}_{i,k}$ sú poporadí výberové priemery prvých k pozorovaní a posledných $T - k$ pozorovaní, oba definované v (3.10).

Za nulovej hypotézy má táto testová štatistika zvyčajne známe asymptotické rozdelenie, ktoré je funkciou vektoru alebo náhodného procesu z rodiny normálnych rozdelení. Okrem neznámych parametrov je však nutné odhadnúť tiež korelačnú štruktúru tohto rozdelenia. Iným prístupom je použitie neparametrického bootstrapu (Peštová a Pešta, 2017).

Najskôr sa odhadnú chybové zložky $\varepsilon_{i,t}$ ako

$$\hat{\varepsilon}_{i,k} := \begin{cases} Y_{i,k} - \bar{Y}_{i,\hat{k}}, & k \leq \hat{k}; \\ Y_{i,k} - \tilde{Y}_{i,\hat{k}}, & k > \hat{k}, \end{cases} \quad (3.12)$$

kde \hat{k} je nejaký konzistentný odhad skutočného bodu zmeny bez okrajovej podmienky.

Vďaka nezávislosti panelov možno korelačnú štruktúru odhadovať empiricky, za predpokladu, že N je dostatočne veľké. Empirický odhad autokorelačnej funkcie, ktorý používajú Peštová a Pešta (2017) má nasledovný tvar

$$\hat{\rho}_k = \frac{1}{N(T-k)} \sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2} \sum_{t=1}^{T-k} \hat{\varepsilon}_{i,t} \hat{\varepsilon}_{i,t+k},$$

kde $\hat{\sigma}_i^2$ je nejaký konzistentný odhad rozptylu σ_i^2 , napríklad $\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{i,t}^2$. Kumulatívna autokorelačná funkcia a posunutá kumulatívna autokorelačná funkcia definované v Predpoklade 10 sa potom odhadnú dosadením odhadu $\hat{\rho}_k$ miesto ρ_k . Andrews (1991) dokázal, že $\hat{\rho}_k$ je konzistentným odhadom ρ_k .

Procedúra, ktorú popisujú Peštová a Pešta (2017), sa zakladá na použití neparametrického bootstrapu na vektory chybových zložiek $\varepsilon_1, \dots, \varepsilon_N$, kde ε_i je riadkový vektor tvaru $\varepsilon_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,T})$ prislúchajúci i -temu panelu. Tieto vektory sú z Predpokladu 10 nezávislé a rovnako rozdelené a empiricky sú odhadnuté pomocou vzťahu (3.12). Následne sa pre B dostatočne veľké a pre každé $b \in \{1, \dots, B\}$ generujú nezávislé dátové súbory $\varepsilon^{(b)} = (\varepsilon_1^{(b)}, \dots, \varepsilon_N^{(b)})$ z empirického rozdelenia (3.12), čo je ekvivalentné náhodnému výberu s vracaním veľkosti N . Získané rezíduá $\hat{\varepsilon}_{i,t}^{(b)}$ sú potom centrované ich podmienenými strednými hodnotami, čo dáva

$$\hat{Y}_{i,t}^{(b)} = \hat{\varepsilon}_{i,t}^{(b)} - \frac{1}{N} \sum_{i=1}^N \hat{\varepsilon}_{i,t}. \quad (3.13)$$

Pôvodné pozorovania $Y_{i,t}$ sa potom v testovej štatistike nahradia vygenerova-

nými pozorovaniami $\widehat{Y}_{i,t}^{(b)}$, čím sa získa

$$S_{N,T}^{(b)} = \max_{k=2,\dots,T-2} \frac{\max_{t=1,\dots,k-1} |L_N^{(b)}(t, k)|}{\max_{t=k,\dots,T-1} |R_N^{(b)}(t, k)|}, \quad (3.14)$$

kde

$$L_N^{(b)}(t, k) := \sum_{i=1}^N \sum_{r=1}^t \left(\widehat{Y}_{i,r}^{(b)} - \overline{\widehat{Y}}_{i,k}^{(b)} \right), \quad R_N^{(b)}(t, k) = \sum_{i=1}^N \sum_{r=t+1}^T \left(\widehat{Y}_{i,r}^{(b)} - \widetilde{\widehat{Y}}_{i,k}^{(b)} \right)$$

a výberové priemery sú definované ako

$$\overline{\widehat{Y}}_{i,k}^{(b)} = \frac{1}{k} \sum_{t=1}^k \widehat{Y}_{i,t}^{(b)}, \quad \widetilde{\widehat{Y}}_{i,k}^{(b)} = \frac{1}{T-k} \sum_{t=k+1}^T \widehat{Y}_{i,t}^{(b)}.$$

Nasledovná veta, ktorú uviedli a dokázali Peštová a Pešta (2017), zaisťuje fungovanie testov založených na uvedenej bootstrapovej procedúre. Okrem testovej štatistiky (3.11) táto veta platí pre ľubovoľnú testovú štatistiku používanú na detekciu bodov zmeny tvaru

$$S_{N,T} \equiv S \left(\left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{r=1}^t (Y_{i,r} - \overline{Y}_{i,k}) \right\}_{t=1,k=2}^{k-1,T}, \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{r=t+1}^T (Y_{i,r} - \widetilde{Y}_{i,k}) \right\}_{t=k,k=1}^{T-1,T-1} \right),$$

kde $S(.,.) : \mathbb{R}^{T \times (T-1)/2} \times \mathbb{R}^{T \times (T-1)/2} \rightarrow \mathbb{R}$ je spojitá funkcia.

Veta 9. *Za Predpokladov 10, 11 a 3. bodu Predpokladu 12,*

1. *za platnosti H_0 ,*

$$S_{N,T} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \mathcal{L};$$

2. *ak navyše platí 1. a 2. bod Predpokladu 12 a za platnosti H_0 , ako aj za platnosti H_1 ,*

$$S_{N,T}^{(b)} | \mathbb{Y} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \mathcal{L}^{(b)}, \quad \text{v pravdepodobnosti } \mathcal{P},$$

kde $\mathbb{Y} \equiv \{Y_{i,t}\}_{i,t=1}^{N,T}$;

3. *ak navyše platí 1. a 2. bod Predpokladu 12 a za platnosti H_0 sú rozdelenia \mathcal{L} a $\mathcal{L}^{(b)}$ totožné v pravdepodobnosti \mathcal{P} .*

Vďaka druhému bodu Vety 9 môže byť bootstrapová testová štatistika $S_{N,T}^{(b)}$ použitá aj v prípadoch, keď nulovú hypotézu o neprítomnosti zmeny zamietame v prospech alternatívy. Testové štatistiky tvaru ako $S_{N,T}$ majú za alternatívy tendenciu divergovať v pravdepodobnosti k nekonečnu. Tretí bod Vety 9 navyše zaručuje, že bootstrapový test zachováva približne rovnakú hladinu spoľahlivosti ako pôvodný test.

3.2.1 Odhad pre závislé dáta

V tejto sekcii si zavedieme odhad bodu zmeny pre závislé dáta, ktorý navrhli Pešta a kol. (2020). Testovanie prítomnosti bodu zmeny pre závislé dáta a konštrukciu bootstrapových testových štatistík vo svojej práci skúmajú Maciak a kol. (2020).

Pešta a kol. (2020) uvažujú α -zmiešané vektory rezíduí. Pre definíciu α -zmiešaných náhodných procesov sa odkazujeme na Doukhan (1994).

Predpoklad 13. Vektory $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,T})$ sú α -zmiešané s nulovou strednou hodnotou, kde koeficienty α splňujú $\sum_{i=1}^{\infty} \{\alpha(\boldsymbol{\varepsilon}_{\bullet}, i)^{\chi/(2+\chi)}\}$, kde pre danú náhodnú postupnosť $\{\boldsymbol{\varepsilon}_n\}_{n \in \mathbb{N}}$ sú koeficienty $\alpha(n)$ označené ako $\alpha(\boldsymbol{\varepsilon}_{\bullet}, n)$, pre nejaké $\chi > 0$ a $\sup_{i \in \mathbb{N}} \mathbf{E} |\varepsilon_{i,t}|^{4+2\chi} < \infty$, $\forall t \in \{1, \dots, T\}$.

Okrem predpokladu nezávislosti panelov je vypustený aj predpoklad o ich rovnakom rozdelení a predpoklad slabej stacionarity. To nám umožňuje pracovať s omnoho širšou triedou rozdelení. Ďalej sa zavádzajú nasledovné predpoklady na vzťahy medzi veľkosťami skokov, počtom panelov a variabilitou rezíduí.

Predpoklad 14.

$$\sup_{i \in \mathbb{N}} |\delta_i| < \infty.$$

Predpoklad 15. Nech $\underline{\varsigma}_i = \min_{u,v \in \{1, \dots, T\}} \mathbf{E} \varepsilon_{i,u} \varepsilon_{i,v}$ a $\bar{\varsigma}_i = \max_{u,v \in \{1, \dots, T\}} \mathbf{E} \varepsilon_{i,u} \varepsilon_{i,v}$. Pre $\kappa < T$ platí

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \left[\left\{ 1 - \frac{2\kappa(T-\kappa)}{(T-1)^2} \right\} \sum_{i=1}^N \delta_i^2 - \frac{2}{T-1} \sum_{i=1}^N (\sigma_i^2 - \underline{\varsigma}_i) \right] = \infty$$

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \sum_{i=1}^N \delta_i^2 = \infty$$

a pre $\kappa = T$ je

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \left(\sum_{i=1}^N (\sigma_i^2 - \bar{\varsigma}_i) - \frac{T-1}{2} \sum_{i=1}^N \delta_i^2 \right) = \infty.$$

Za týchto predpokladov dostaneme konzistentný odhad $\tilde{\kappa}_P$, ktorý je definovaný ako

$$\tilde{\kappa}_P := \arg \max_{1 \leq k \leq T} \frac{1}{k(T-k)} \sum_{i=1}^N \sum_{t=1}^k \sum_{u=k+1}^T (Y_{i,t} - Y_{i,u})^2, \quad k < T$$

a pre $k = T$ sa priradí

$$\tilde{\kappa}_P := \arg \max_{1 \leq k \leq T} \frac{2}{(T-1)^2} \sum_{i=1}^N \sum_{u=2}^T \sum_{t=1}^{u-1} (Y_{i,t} - Y_{i,u})^2.$$

Veta 10. Za Predpokladov 13 – 15 platí

$$\lim_{N \rightarrow \infty} P(\tilde{\kappa}_P = \kappa) = 1.$$

Ďalej chceme testovať nulovú hypotézu o neprítomnosti bodu zmeny v jednotlivých paneloch.

Predpoklad 16. Vektory $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,T})$ sú α -zmiešané s nulovou strednou hodnotou, kde koeficienty α splňujú $\sum_{i=1}^{\infty} \{\alpha(\varepsilon_{\bullet, i})^{\chi/(2+\chi)}\}$ pre nejaké $\chi > 0$ a $\sup_{i \in \mathbb{N}} \mathbf{E} |\varepsilon_{i,t}|^{2+\chi} < \infty$, $\forall t \in \{1, \dots, T\}$. Navyiac existuje pozitívne definitná matica

$$\boldsymbol{\Lambda}_N := \lim_{N \rightarrow \infty} \frac{1}{N} \text{var} \left(\sum_{i=1}^N \left[\varepsilon_{i,1}, \sum_{s=1}^2 \varepsilon_{i,s}, \dots, \sum_{s=1}^T \varepsilon_{i,s} \right]^T \right).$$

Maciak a kol. (2020) využívajú na testovanie hypotézy (3.2) normalizované testové štatistiky tvaru

$$Q_{N,T} = \max_{k=1, \dots, T-1} \frac{|L_N(k, T)|}{\max_{t=1, \dots, k} |L_N(t, k)| + \max_{t=k, \dots, T-1} |R_N(t, k)|} \quad (3.15)$$

a

$$K_{N,T} = \sum_{k=1}^{T-1} \frac{L_N^2(k, T)}{\sum_{t=1}^k L_N^2(t, k) + \sum_{t=k}^{T-1} R_N^2(t, k)} \quad (3.16)$$

kde $L_N(t, k)$ a $R_N(t, k)$ sú definované v (3.11). Výhodou týchto testových štatistik je, že nezávisia na rozptyle chybových zložiek, a teda tento parameter nie je nutné odhadovať.

Nasledujúca veta hovorí o asymptotickom správaní týchto testových štatistik za platnosti nulovej hypotézy.

Veta 11. *Za Predpokladu 16 a platnosti nulovej hypotézy platí*

$$Q_{N,T} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \max_{k=1, \dots, T-1} \frac{|X_k - \frac{k}{T} X_T|}{\max_{t=1, \dots, k} |X_t - \frac{t}{k} X_k| + \max_{t=k, \dots, T-1} |Z_t - \frac{T-t}{T-k} Z_k|} \quad (3.17)$$

a

$$K_{N,T} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \sum_{k=1}^{T-1} \frac{(X_k - \frac{k}{T} X_T)^2}{\sum_{t=1}^k (X_t - \frac{t}{k} X_k)^2 + \sum_{t=k}^{T-1} (Z_t - \frac{T-t}{T-k} Z_k)^2}, \quad (3.18)$$

kde $Z_k := X_T - X_k$ a vektor $[X_1, \dots, X_T]^T$ má mnohorozmerné normálne rozdelenie s nulovou strednou hodnotou a variančnou maticou $\boldsymbol{\Lambda}_N$.

Maciak a kol. (2020) ďalej uvažujú nasledovný predpoklad na veľkosti skokov v jednotlivých paneloch. Tento predpoklad potom používajú pri odvodzovaní asymptotického rozdelenia testových štatistik (3.17) a (3.18) za platnosti alternatívy.

Predpoklad 17.

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \left| \sum_{i=1}^N \delta_i \right| = \infty.$$

Za takýchto podmienok divergujú obe testové štatistiky za platnosti alternatívy v pravdepodobnosti do nekonečna.

Veta 12. *Za Predpokladov 16, 17 a za platnosti alternatívy platí*

$$Q_{N,T} \xrightarrow[N \rightarrow \infty]{P} \infty$$

$$K_{N,T} \xrightarrow[N \rightarrow \infty]{P} \infty.$$

Keďže v Predpoklade 17 vystupujú absolútne hodnoty, existuje tu možnosť vyrušenia skokov s kladnými a zápornými znamienkami, čím by Predpoklad zrejme nemohol byť splnený. Maciak a kol. (2020) toto ospravedľujú faktom, že v reálnych aplikáciách majú zmeny, ak sú v dátach nejaké prítomné, tendenciu nadobúdať rovnaké znamienka. Za opačných podmienok by však porušenie nastalo samozrejme iba v prípade, že by bol počet kladných a záporných zmien rovnaký. Existuje teda široká trieda parametrov δ_i , na ktorú možno odvodené teoretické poznatky aplikovať.

Z Vety 12 dostávame, že navrhované procedúry sú konzistentné a asymptotické rozdelenia vo Vete 11 môžu byť použité na konštrukciu štatistických testov. Maciak a kol. (2020) konštruujú testy bez rušivých parametrov, ktoré odhadujú pomocou neparametrického bootstrapu.

Bootstrapové testové štatistiky $Q_{N,T}^{(b)}$ a $K_{N,T}^{(b)}$ vzniknú nahradením pozorovaní $Y_{i,t}$ za bootstrapové pozorovania $\hat{Y}_{i,t}^{(b)}$ v pôvodných testových štatistikách.

Maciak a kol. (2020) podrobne vysvetľujú pomocou nasledovného algoritmu fungovanie bootstrapovej procedúry s takzvanými pohyblivými blokmi riadkových vektorov rezíduí. Pre jednoduchosť predpokladajú, že $N = bn$, kde b je veľkosť bloku a n je ich počet. Ak by pre počet panelov platilo, že $N = bn + n_0$ pre nejaké $1 \leq n_0 \leq n - 1$, limitné výsledky by sa po vynechaní zvyšných n_0 pozorovaní nezmenili (Maciak a kol., 2020).

Algoritmus fungovania bootstrapu s pohyblivými blokmi riadkových vektorov rezíduí

Vstup: blok veľkosti b , počet bootstrapových replikácií M , panelové dáta pozostávajúce z $N = bn$ panelov dĺžky T

Výstup: bootstrapové rozdelenia $Q_{N,T}^{(b)}$ a $K_{N,T}^{(b)}$, tj. empirické rozdelenie, kde ${}_{(1)}Q_{N,T}^{(b)}, \dots, {}_{(M)}Q_{N,T}^{(b)}$ a ${}_{(1)}K_{N,T}^{(b)}, \dots, {}_{(M)}K_{N,T}^{(b)}$ sa nadobúdajú s rovnakou pravdepodobnosťou $1/M$

1. Odhadni bod zmeny ako $\hat{\kappa}$.
2. Vypočítaj odhady $\hat{\epsilon}_{i,k}$ pomocou vzorca (3.12).
3. Pre $j = 1, \dots, N - b + 1$ skonštruuj bloky:
 - definuj blok podpanelov \mathbb{B}_j ako maticu typu $(b \times T)$ s riadkami $\hat{\epsilon}_j, \dots, \hat{\epsilon}_{j+b-1}$.
4. Pre $m = 1, \dots, M$ opakuj kroky
 - vykonaj nezávislý náhodný výber s vracaním $(b \times T)$ -rozmerných blokov ${}_{(m)}\mathbb{B}_1^{(b)}, \dots, {}_{(m)}\mathbb{B}_n^{(b)}$ z rozdelenia $\{\mathbb{B}_1, \dots, \mathbb{B}_{N-b+1}\}$ s rovnakou pravdepodobnosťou výberu $1/(N - b + 1)$;

- bootstrapový pohyblivý blok výberu veľkosti N je značený ako $\{({}_{(m)}\hat{\varepsilon}_{i,k}^{(b)})\}_{i,k=1}^{N,T}$ a je vytvorený združením vygenerovaných blokov $({}_{(m)}\mathbb{B}_1^{(b)}, \dots, {}_{(m)}\mathbb{B}_n^{(b)})$ do jedného bloku $[({}_{(m)}\mathbb{B}_1^{(b)T}, \dots, {}_{(m)}\mathbb{B}_n^{(b)T})^T = [({}_{(m)}\hat{\varepsilon}_1^{(b)T}, \dots, {}_{(m)}\hat{\varepsilon}_n^{(b)T})^T]$;
- vypočítaj $({}_{(m)}\hat{Y}_{i,t}^{(b)})$ pomocou (3.13);
- vypočítaj bootstrapové testové štatistiky $({}_{(m)}Q_{N,T}^{(b)})$ a $({}_{(m)}K_{N,T}^{(b)})$ ako v (3.15) a (3.16), kde $Y_{i,k}$ sú nahradené $\hat{Y}_{i,k}^{(b)}$.

Veľkosť blokov b je nutné vhodne zvoliť, keďže ovplyvňuje výsledné bootstrapové testové štatistiky. Jedným z prístupov je testovať rôzne hodnoty b pri optimalizácii nejakého vhodného kritéria.

Na to, aby sa ukázalo, že uvedenú procedúru je možné používať ku konštrukcii testov, (Maciak a kol., 2020) formulujú silnejší predpoklad na chybové zložky modelu než je Predpoklad 16.

Predpoklad 18. Vektory $\varepsilon_i \equiv [\varepsilon_{i,1}, \dots, \varepsilon_{i,T}]^T$ sú α -zmiešané s nulovou strednou hodnotou, kde koeficienty α spĺňujú $\alpha(\varepsilon_{\bullet}, N) = O(N^{-1-\bar{\omega}})$ pre $N \rightarrow \infty$ a nejaké $\bar{\omega} > 0$ a $\sup_{i \in \mathbb{N}} \mathbf{E} |\varepsilon_{i,k}|^{4+\omega} < \infty$ pre $\forall k \in \{1, \dots, T\}$ a $\omega > 0$ také, že $\omega\bar{\omega} > 4$. Navyše existuje pozitívne definitná matica

$$\mathbf{\Gamma}_N := \lim_{N \rightarrow \infty} \frac{1}{N} \text{var} \sum_{i=1}^N \left[e_{i,1}, \sum_{t=1}^2 e_{i,t}, \dots, \sum_{t=1}^T e_{i,t} \right]^T,$$

kde

$$e_{i,k} = \begin{cases} \varepsilon_{i,k} - \bar{\varepsilon}_{i,\kappa}, & k \leq \kappa; \\ \varepsilon_{i,k} - \tilde{\varepsilon}_{i,\kappa}, & k > \kappa, \end{cases} \quad \bar{\varepsilon}_{i,k} = \frac{1}{k} \sum_{t=1}^k \varepsilon_{i,t}, \quad \text{a} \quad \tilde{\varepsilon}_{i,k} = \frac{1}{T-k} \sum_{t=k+1}^T \varepsilon_{i,t}.$$

Ďalej je potrebný konzistentný odhad bodu zmeny.

Predpoklad 19. $\lim_{N \rightarrow \infty} \mathbf{P}(\hat{\kappa} = \kappa) = 1$ pre $\kappa \in \{1, \dots, T\}$.

Nasledovná veta ukazuje, že rozdelenia podmienených bootstrapových testových štatistik konvergujú k rozdeleniu, ktoré je funkciou mnohorozmerného normálneho vektoru s nulovou strednou hodnotou a známou variančnou štruktúrou. Uviedli ju Maciak a kol. (2020).

Veta 13. *Nech platia predpoklady 18 a 19. Ak platí $b \rightarrow \infty$ a $b = o(N^{1/2})$ pre $N \rightarrow \infty$, potom*

$$Q_{N,T}^{(b)} | \mathbb{Y} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \max_{k=1, \dots, T-1} \frac{|\mathcal{X}_k - \frac{k}{T} \mathcal{X}_T|}{\max_{t=1, \dots, k} |\mathcal{X}_t - \frac{t}{k} \mathcal{X}_k| + \max_{t=k, \dots, T-1} |\mathcal{Z}_t - \frac{T-t}{T-k} \mathcal{Z}_k|}$$

a

$$K_{N,T}^{(b)} | \mathbb{Y} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \sum_{k=1}^{T-1} \frac{(\mathcal{X}_k - \frac{k}{T} \mathcal{X}_T)^2}{\sum_{t=1}^k (\mathcal{X}_t - \frac{t}{k} \mathcal{X}_k)^2 + \sum_{t=k}^{T-1} (\mathcal{Z}_t - \frac{T-t}{T-k} \mathcal{Z}_k)^2}$$

v pravdepodobnosti, kde $\mathcal{Z}_k := \mathcal{X}_T - \mathcal{X}_k$ a vektor $[\mathcal{X}_1, \dots, \mathcal{X}_T]^T$ má mnohorozmerné normálne rozdelenie s nulovou strednou hodnotou a variančnou maticou $\mathbf{\Gamma}_N$.

Veta 13 platí za H_0 aj za H_1 . Podobne ako aj v prípade závislých dát je teda pri dostatočnom počte panelov N bootstrap validný pri zamietaní hypotézy v prospech alternatívy, keďže jeho limitné rozdelenie neexploduje za alternatívy do nekonečna.

Označme ďalej $\mathcal{L}(W)$ limitné rozdelenie W . Maciak a kol. (2020) uvádzajú vetu o konzistencii popísanej bootstrapovej procedúry.

Veta 14. *Nech platia Predpoklady 18, 19 a hypotéza H_0 . Ak $b \rightarrow \infty$ a $b = O(N^{1/2})$ pre $N \rightarrow \infty$, potom*

i. limitné rozdelenia $\mathcal{L}(Q_{N,T}^{(b)}|\mathbb{Y})$ a $\mathcal{L}(Q_{N,T})$ sa zhodujú v pravdepodobnosti P ;

ii. limitné rozdelenia $\mathcal{L}(K_{N,T}^{(b)}|\mathbb{Y})$ a $\mathcal{L}(K_{N,T})$ sa zhodujú v pravdepodobnosti P .

Z Vety 14 plynie, že bootstrapové testy budú dodržiavať približne rovnakú hladinu ako testy založené na pôvodných testových štatistikách. Je ich teda možné použiť ku konštrukcii testov porovnaním bootstrapových kritických hodnôt s kritickými hodnotami pôvodných štatistik.

3.3 Odhad bodu zmeny pomocou kvazi-vierohodnostnej funkcie

Horváth a Hušková (2012) sa vo svojom článku zaoberajú odhadovaním bodu zmeny v panelových dátach za použitia kvazi-vierohodnostnej funkcie. Zámerom je testovať nulovú hypotézu (3.2) o nemennej strednej hodnote oproti alternatíve, že stredná hodnota sa aspoň raz za sledované obdobie zmení. Vychádza sa z modelu (3.1) pre T a N dostatočne veľké.

S touto metódou prišiel už Bai (2010), odkiaľ tiež autori článku vychádzajú. Nakoľko sa uvažuje $T \rightarrow \infty$, o bode zmeny sa predpokladá, že je tvaru $\kappa = \lfloor Tx \rfloor$, $x \in [0, 1]$.

Horváth a Hušková (2012) k testovaniu H_0 využívajú testovú štatistiku tvaru

$$\sup_{0 \leq x \leq 1} |\bar{V}_{N,T}(x)|,$$

kde

$$\bar{V}_{N,T}(x) = \frac{1}{N^{1/2}} \sum_{i=1}^N \left\{ \frac{1}{\omega_i} Z_{T,i}^2(x) - \frac{\lfloor Tx \rfloor (T - \lfloor Tx \rfloor)}{T^2} \right\}, \quad x \in [0, 1],$$

$Z_{T,i}$ je podobne ako vo vyjadrení (1.5) pre jeden panel definované ako

$$Z_{T,i}(x) = \frac{1}{\sqrt{T}} \left(\sum_{t=1}^{\lfloor Tx \rfloor} Y_{i,t} - \frac{\lfloor Tx \rfloor}{T} \sum_{t=1}^T Y_{i,t} \right), \quad x \in [0, 1]$$

a nakoniec ω_i je definované ako

$$\omega_i = \lim_{T \rightarrow \infty} \frac{1}{T} \text{var} \left(\sum_{t=1}^T \varepsilon_{i,t} \right), \quad 1 \leq i \leq N,$$

teda ide o *long - run variance* i -teho panelu.

Odhad $\hat{\kappa}_{QL}$ bodu κ pomocou kvazi-vierohodnostnej funkcie je definovaný ako

$$\hat{\kappa}_{QL} = \arg \max_{1 \leq k \leq T-1} |\bar{V}_{N,T}(k/T)|.$$

Aj tu sa predpokladá, že $\{\varepsilon_{i,t}\}_t$ tvoria lineárny proces a sú nezávislé medzi panelmi. (viď Predpoklad 4). Ďalej sa uvažujú jednotkové rozptyly v definícii lineárneho procesu, tj. $\mathbf{E}(e_{i,t})^2 = 1$, keďže testová štatistika $\bar{V}_{N,T}$ nezávisí na reziduálnom rozptyle. Navyše sa predpokladá, že ξ -te absolútne momenty veličín $e_{i,t}$ sú konečné

$$\mathbf{E} |e_{i,0}|^\xi < \infty.$$

Za ξ sa berie $\xi > 4$ alebo $\xi = 8$. Ďalšou podmienkou je podmienka na ohraničenosť priemeru vyšších momentov

$$\lim_{N \rightarrow \infty} \sup \frac{1}{N} \sum_{i=1}^N \mathbf{E} |e_{i,0}|^\xi < \infty. \quad (3.19)$$

Silnejšou podmienkou než je (3.19) sú rovnomerne ohraničené momenty $\mathbf{E} |e_{i,0}|^\xi$. Podmienka (3.19) však umožňuje, aby boli niektoré momenty $\mathbf{E} |e_{i,0}|^\xi$, a teda aj momenty chýb $\mathbf{E} |\varepsilon_{i,0}|^\xi$ veľké (Horváth a Hušková, 2012).

Z vlastností koeficientov v definícii lineárneho procesu a z predošlých predpokladov sa ďalej odvodí podmienka na rovnomernú ohraničenosť rozptylov ω_i medzi panelmi zdola

$$\omega_i \geq \sigma^2 \quad \forall 1 \leq i \leq N$$

pre nejaké $\sigma > 0$.

Na rozdiel od Predpokladu 6, ktorý vyžaduje, aby bol počet pozorovaní T v každom paneli vyšší než je počet panelov N , tu sa uvažuje nasledovný predpoklad, ktorý umožňuje, aby bolo N väčšie než je T .

Predpoklad 20.

$$\lim_{N,T \rightarrow \infty} \frac{N}{T^2} = 0.$$

Horváth a Hušková (2012) ukazujú, že za takýchto podmienok testová štatistika $\bar{V}_{N,T}(x)$ slabo konverguje v Skorokhodovom priestore $\mathcal{D}[0, 1]$, ktorý definuje aj Micheas (2018), ku Gaussovskému procesu $\Gamma(x)$ s nulovou strednou hodnotou a známou kovariančnou štruktúrou. Z toho potom plynie, že testová štatistika $\sup_{0 \leq x \leq 1} |\bar{V}_{N,T}(x)|$ konverguje v distribúcii k náhodnému procesu $\sup_{0 \leq x \leq 1} |\Gamma(x)|$. Veľké hodnoty tejto testovej štatistiky potom svedčia proti H_0 . Túto problematiku podrobne vysvetľujú Horváth a Hušková (2012) a tiež stanovujú podmienky pre konzistenciu testov. Test je citlivý na väčšie zmeny vyskytujúce sa v relatívne malom množstve panelov a tiež je citlivý na relatívne malé zmeny vyskytujúce sa vo veľkom počte panelov. Horváth a Hušková (2012) taktiež poskytujú asymptotické kritické hodnoty testovej štatistiky $\sup_{0 \leq x \leq 1} |\bar{V}_{N,T}(x)|$, získané simuláciami.

3.3.1 Odhad rozptylov

Za predpokladu, že $\{\varepsilon_{i,t}\}_t$ sú v každom paneli i nezávislé a rovnako rozdelené, Horváth a Hušková (2012) používajú na odhadovanie parametru ω_i výberový rozptyl

$$\hat{\omega}_{i,T} = \frac{1}{T-1} \sum_{t=1}^T (Y_{i,t} - \bar{Y}_{i,T})^2, \quad \bar{Y}_{i,T} = \frac{1}{T} \sum_{t=1}^T Y_{i,t}. \quad (3.20)$$

Pre $\{\varepsilon_{i,t}\}_t$ závislé využívajú jadrový odhad

$$\hat{\omega}_{i,T} = \frac{1}{T} \sum_{t=1}^T (Y_{i,t} - \bar{Y}_{i,T})^2 + 2 \sum_{s=1}^{T-1} K\left(\frac{s}{h}\right) \hat{\gamma}_{i,T}(s), \quad (3.21)$$

kde

$$\hat{\gamma}_{i,T}(s) = \frac{1}{T-s} \sum_{t=1}^{T-s} (Y_{i,t} - \bar{Y}_{i,T})(Y_{i,t+s} - \bar{Y}_{i,T})$$

je výberová kovariancia rádu s pre i -ty panel, funkcia K je jadro a $h = h(T)$ je vyhladzovací parameter.

Označme ďalej, podobne ako Horváth a Hušková (2012),

$$V_{N,T}(x) = \frac{1}{N^{1/2}} \sum_{i=1}^N \left\{ \frac{1}{\hat{\omega}_{i,T}} Z_{T,i}^2(x) - \frac{\kappa(T-\kappa)}{T^2} \right\}, \quad x \in [0, 1].$$

Aby sa dokázala slabá konzistencia testovej štatistiky $V_{N,T}$, zavádzajú sa ďalšie podmienky na jadro a vyhladzovací parameter (podrobnejšie viď Horváth a Hušková (2012)). Za týchto podmienok potom $V_{N,T}$ slabo konverguje k procesu $\Gamma(x)$. Takúto testovú štatistiku je potom možné použiť ku konštrukcii testov.

3.3.2 Závislé panely

Horváth a Hušková (2012) vnášajú do modelu (3.1) závislosť medzi panelmi prostredníctvom premennej ν_t

$$Y_{i,t} = \mu_i + \delta_i I[t > \kappa] + \phi_i \nu_t + \varepsilon_{i,t}, \quad 1 \leq i \leq N, 1 \leq t \leq T, \quad (3.22)$$

kde

$$\{\nu_t, 1 \leq t \leq \infty\} \text{ a } \{\varepsilon_{i,t}, 1 \leq t \leq \infty\}$$

sú nezávislé. Parametre ν_t majú navyše splňovať

$$\frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor Tx \rfloor} \nu_t \xrightarrow[T \rightarrow \infty]{\mathcal{D}[0,1]} W(x),$$

kde $W(x)$ je Wienerov proces a $\xrightarrow[T \rightarrow \infty]{\mathcal{D}[0,1]}$ značí slabú konvergenciu v Skorokhodovom priestore $\mathcal{D}[0, 1]$. Konštanty ϕ_i udávajú, aký veľký vplyv majú parametre ν_t na i -ty panel. Horváth a Hušková (2012) ukazujú, že aj malá závislosť medzi panelmi mení asymptotické správanie testových štatistík $\bar{V}_{N,T}$ a $V_{N,T}$ za H_0 . Ak je však korelácia medzi panelmi menšia než $N^{-1/2}$, potom závislosť na testové štatistiky nemá vplyv. Podrobnejšie vysvetlenie sa nachádza v článku Horváth a Hušková (2012).

4. Analýza reálnych dát

V tejto kapitole aplikujeme metódu WBS popísanú v Kapitole 2 na reálne dáta. Budeme skúmať vplyv rôznych penalizačných kritérií WBS metódy na odhadnutý počet a polohy bodov zmeny.

4.1 Dátový súbor

Analyzujeme týždenné logaritmické výnosy záverečného kurzu Dow Jones indexu v období od 01.04.1990 až 01.01.2012. Dokopy máme k dispozícii $T = 1139$ pozorovaní. Použitý dátový súbor analyzovali James a Matteson (2014). V skúmanom období sa udiali nasledovné udalosti, pri ktorých očakávame výskyt zmien v strednej hodnote analyzovaného akciového indexu:

1. Recesia na počiatku 90.-tych rokov minulého storočia – začala v júli 1990 a trvala približne 8 mesiacov;
2. Ázijská finančná kríza v roku 1997 – začala v júli 1997 a vyústila v globálny krach na akciovej burze v októbri 1997;
3. Ruská finančná kríza v roku 1998 – začiatok v auguste 1998;
4. Útoky 11. septembra 2001;
5. Pokles cien na akciovom trhu v roku 2002 – najväčšie prepady cien akcií sa udiali v júli až septembri a v októbri, pričom posledný väčší pokles nastal v marci 2003;
6. Finančná kríza 2007–2009 – najväčšie prepady Dow Jones indexu v septembri 2008 a marci 2009;
7. Prepad cien na akciovom trhu v auguste 2011.

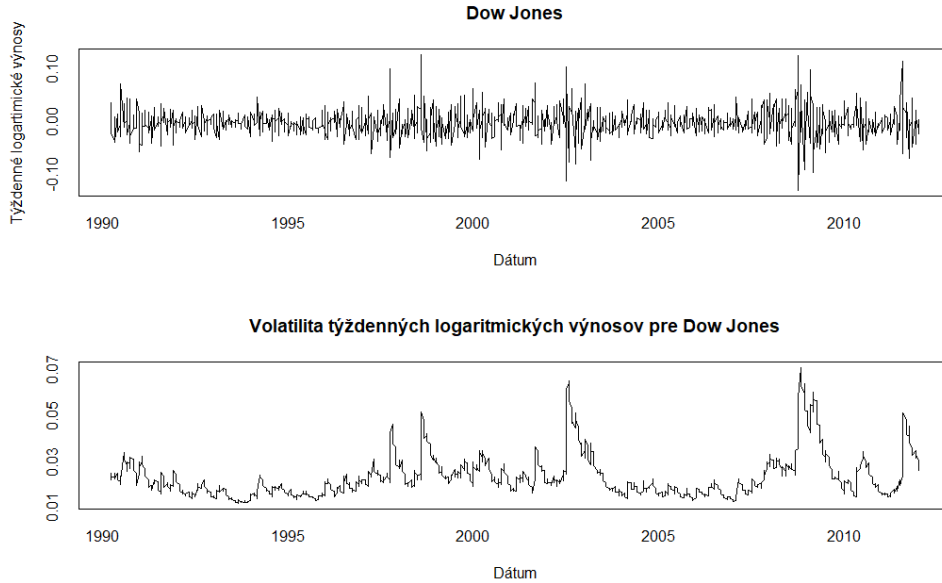
4.2 Popis modelov

Metódu WBS najskôr aplikujeme na pôvodný rad týždenných logaritmických výnosov Dow Jones. Model s pôvodným radom budeme ďalej značiť ako model DJI. Zapísať ho môžeme nasledovne.

$$Y_t = f_t + \varepsilon_t, \quad t = 1, \dots, 1139, \quad (4.1)$$

kde Y_t , je týždenný logaritmický kurz Dow Jones v čase t , f_t je po častiach konštantná stredná hodnota radu Y_t s bodmi zmeny $0 = \kappa_0 < \kappa_1 < \dots < \kappa_m < \kappa_{m+1} = T = 1139$ a $\{\varepsilon_t\}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

Na Obrázku 4.1 týždenných logaritmických výnosov Dow Jones indexu (horný graf) môžeme vidieť, že rad vykazuje zhlukovanie volatility v niekoľkých úsekoch.



Obr. 4.1: Týždenné logaritmické výnosy Dow Jones indexu a volatilita odhadnutá pomocou GARCH(1,1) modelu s normálnymi inováciami.

Volatilita je vyššia najmä počas obdobia finančnej krízy, či v období po udalostiach v septembri 2001. Časové rady s takýmito rysmi sa typicky modelujú pomocou nelineárnych modelov, pričom jedným z takýchto modelov je aj model GARCH. Model GARCH spolu s ďalšími nelineárnymi modelmi vysvetľuje vo svojej knihe Cipra (2008). Na rad aplikujeme GARCH(1,1) model s inováciami so štandardným normálnym rozdelením, ktorý má nasledovný tvar

$$Y_t = \mu_t + \tilde{\varepsilon}_t, \quad \tilde{\varepsilon}_t = \sigma_t e_t, \quad t = 1, \dots, 1139, \quad (4.2)$$

kde Y_t je ako vyššie, μ_t je stredná hodnota radu, $\{e_t\}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ a

$$\sigma_t^2 = \alpha_0 + \alpha_1 \tilde{\varepsilon}_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad \alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1.$$

Parameter	Odhad	Št. odchýlka	Test. št.	p-hod.
μ_t	-0,002	0,000	-4,015	< 0,001
α_0	0,000	0,000	2,810	0,005
α_1	0,132	0,023	5,681	< 0,001
β_1	0,847	0,026	33,175	< 0,001

Tabuľka 4.1: Odhady koeficientov GARCH(1,1) modelu so štandardnými normálnymi inováciami, odhady štandardných odchýlok a hodnoty testových štatistík t-testov s príslušnými p-hodnotami.

Na odhadovanie GARCH modelu používame R balíček `fGarch` od Wuertz, D. (2020). Tabuľka 4.1 obsahuje odhady parametrov modelu (4.2). Testové štatistiky prislúchajúce t-testom pre jednotlivé koeficienty modelu majú nízke hodnoty a všetky koeficienty vyšli významné.

Obrázok 4.1 ďalej znázorňuje volatilitu (tj. podmienený rozptyl σ_t^2) skúmaného radu výnosov Dow Jones (dolný graf), ktorá bola odhadnutá pomocou modelu (4.2). V ďalšej časti aplikujeme WBS metódu na odhadnuté rezíduá $\hat{\varepsilon}_t$ z modelu (4.2), teda budeme uvažovať model (4.1) s $Y_t = \hat{\varepsilon}_t$. Na tento model sa budeme odkazovať ako na model GARCH.

4.3 Aplikácia WBS metódy na dáta

V tejto časti použijeme WBS metódu na detekciu bodov zmeny v strednej hodnote týždenných logaritmických výnosov Dow Jones indexu. WBS metódu najskôr aplikujeme na pôvodný rad výnosov a následne na odhadnuté rezíduá modelu GARCH(1,1). Použité funkcie sú predprogramované v R balíčku `breakfast`, ktorého autorom je Fryzlewicz, P. (2017).

Budeme skúmať vplyv penalizačných kritérií na odhadnutý počet a polohy bodov zmeny. Tiež sa pozrieme, aký vplyv na odhadnuté body zmeny malo použitie GARCH(1,1) modelu s normálnymi inováciami. Do analýzy zahrnieme štyri voľby penalizačných kritérií s nasledovnými nastaveniami parametrov:

1. WBS sSIC $\alpha = 1$ – ako penalizačné kritérium je použité Schwarzovo informačné kritérium;
2. WBS $C = 1$ – pevné nastavenie penalizačného parametra ζ_T s $C = 1$;
3. WBS $C = 1,3$ – pevné nastavenie penalizačného parametra ζ_T s $C = 1,3$;
4. WBS auto – automatické nastavenie parametrov C a G tak, aby pravdepodobnosť, že metóda nedetekuje žiadny bod zmeny v konštantnom signáli bez bodov zmeny, bola $\lambda = 0,95$.

Počet vygenerovaných náhodných intervalov je v celej analýze nastavený na $G = 10000$ (až na WBS auto; v balíčku `breakfast` je počet náhodných intervalov značený M). Fryzlewicz (2014) použil pre dátový súbor dĺžky $T = 2000$ hodnotu $G = 5000$. S touto voľbou G dosiahol dostatočne nízku závislosť výsledku na danom náhodnom výbere. S rastúcim G sa však samozrejme dosiahnu presnejšie výsledky. Pri dĺžke nášho dátového súboru sa zdá byť voľba $G = 10000$ postačujúca. Voľba parametra C je motivovaná v Kapitole 2. Ako odhad parametra σ používame prednastavený odhad mediánovej absolútnej deviácie, ktorá sa počíta za predpokladu normálneho rozdelenia reziduálnej zložky.

Každá z uvedených metód bola na dátovom súbore spustená 100-krát a následne boli z odhadnutých modelov napočítané popisné charakteristiky odhadnutého počtu bodov zmeny \hat{m} . Zhrnuté sú v Tabuľkách 4.2 a 4.3. Z popisných charakteristík zrejme môžeme usúdiť, že všetky metódy odhadovali približne rovnaký počet bodov zmeny pre pôvodný rad výnosov ako pre rezíduá z GARCH modelu. V priemere najvyšší počet bodov zmeny \hat{m} odhadovala WBS metóda s $C = 1$. Najnižší počet bodov zmeny zaznamenala WBS metóda s automatickou voľbou parametrov pre oba modely.

Zaujímavejšie môžu byť odhadnuté polohy bodov zmeny (získané z jednej iterácie algoritmu) zhrnuté v Tabuľkách 4.4 až 4.8. Graficky sú znázornené na

Metóda	Model	Min.	I.QR	Výb. medián
WBS sSIC $\alpha = 1$	DJI	0,00	7,00	9,00
WBS $C = 1,0$		13,00	15,75	17,00
WBS $C = 1,3$		9,00	11,00	12,00
WBS auto ¹		3,00	6,00	7,00
WBS sSIC $\alpha = 1$	GARCH	0,00	7,00	9,00
WBS $C = 1,0$		14,00	16,00	16,00
WBS $C = 1,3$		9,00	11,00	12,00
WBS auto		4,00	6,00	7,00

Pozn: ¹ WBS metóda s automatickým nastavením parametrov C a G .

Tabuľka 4.2: Popisné charakteristiky odhadnutého počtu bodov zmeny \hat{m} napočítané zo 100 prevedení metód.

Metóda	Model	Výb.priemer	III.QR	Max.
WBS sSIC $\alpha = 1$	DJI	8,31	11,00	13,00
WBS $C = 1,0$		16,52	17,00	22,00
WBS $C = 1,3$		12,10	13,00	15,00
WBS auto		6,76	7,25	10,00
WBS sSIC $\alpha = 1$	GARCH(1,1)	7,78	11,00	14,00
WBS $C = 1,0$		16,63	17,00	20,00
WBS $C = 1,3$		12,08	13,00	15,00
WBS auto		6,79	8,00	10,00

Tabuľka 4.3: Popisné charakteristiky odhadnutého počtu bodov zmeny \hat{m} napočítané zo 100 prevedení metód.

Obrázku 4.2 pre pôvodný rad týždenných logaritmických výnosov Dow Jones a na Obrázku 4.3 pre rezíduá z modelu GARCH. Polohy sú vyobrazené s mesačnou frekvenciou v sledovanom období.

Metódy WBS sSIC a WBS auto odhadovali od začiatku sledovaného obdobia až do roku 2009 v jednotlivých mesiacoch rovnaké počty bodov zmeny pre pôvodný rad aj pre GARCH model. Najväčšie rozdiely v odhadnutých polohách medzi pôvodným radom a GARCH rezíduami sme zaznamenali pri WBS metóde s ručnou voľbou parametrov C a G .

V januári 1991 metódy celkovo našli jeden bod zmeny odhadnutý pomocou WBS metódy s $C = 1$ pre pôvodný rad (tj. model DJI), konkrétne 07.01.1991. Bod zmeny môže súvisieť s recesiou na začiatku 90.-tych rokov, ktorá trvala približne do marca 1991. V tomto mesiaci sme zaznamenali pomerne prudký pokles kumulatívneho súčtu týždenných výnosov (viď Obrázok 4.4). Celkovo 6 bodov zmeny bolo metódami odhadnutých v októbri 1997, pričom všetky boli odhadnuté pomocou WBS s ručnou voľbou parametrov. Predpokladáme, že tieto zmeny súvisia s krachom na akciovej burze v tom istom mesiaci. Odhadnuté boli dva rôzne body zmeny, a to 13.10.1997 a 20.10.1997. Veľké dopady na výnosy akci-

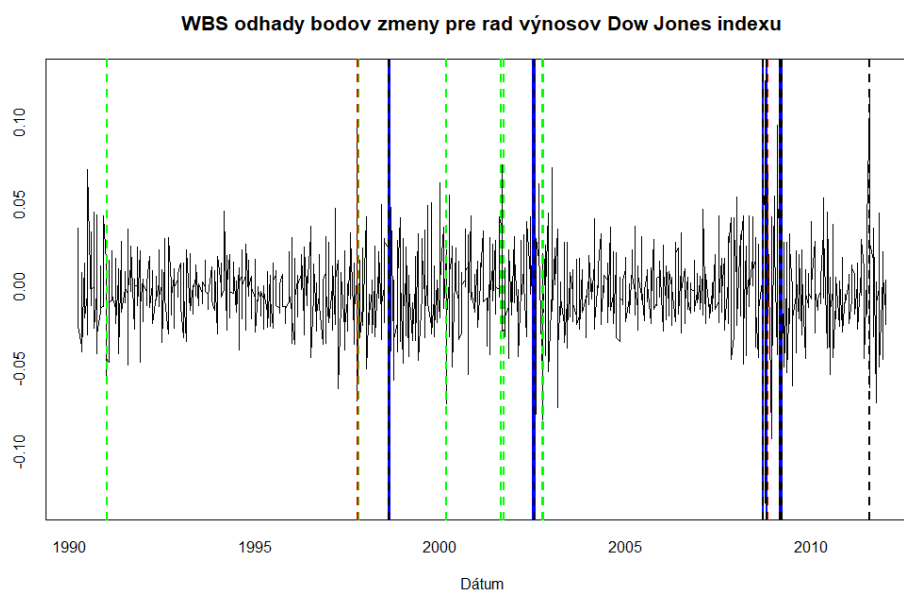
ového indexu mala podľa našich modelov tiež ruská finančná kríza, ktorá začala v auguste 1998. V tomto mesiaci detekujú bod zmeny už všetky skúmané metódy. Celkovo bolo detekovaných 14 bodov zmeny, všetky pripadli na dátumy 17.08.1998 a 24.08.1998.

Ďalšie body zmeny v marci 2000 až septembri 2001 boli opäť detekované pomocou WBS s ručnou voľbou parametrov. Konkrétne boli detekované dátumy 06.03.2000, 20.08.2001 a 17.09.2001. Bod detekovaný v septembri 2001 pravdepodobne súvisí s útokmi, ktoré sa udiali 11.09.2001.

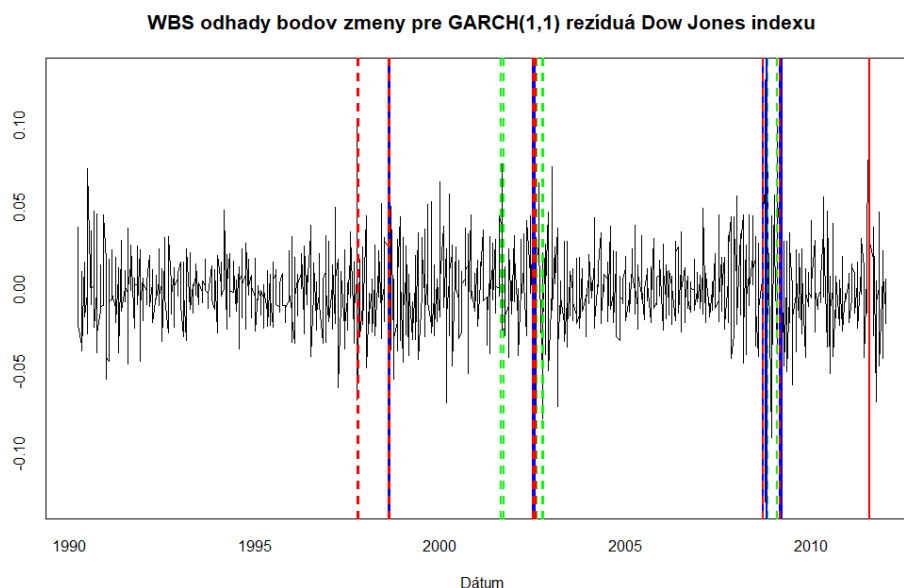
Najväčší počet bodov zmeny bol detekovaný v júli 2002, celkovo bolo nájdených 21 bodov zmeny. Júl 2002 je tiež mesiac s najvyššou frekvenciou bodov zmeny za mesiac. Dátumy, na ktoré body zmeny pripadli, sú 01.07.2002, 15.07.2002, 22.07.2002 a 29.07.2002. Najčastejšie bol detekovaný dátum 15.07.2002, ktorý bol ako jediný detekovaný WBS s automatickou voľbou parametrov. Dátum 29.07.2002 bol odhadnutý iba WBS s parametrom $C = 1,3$ pre model GARCH. Ďalšie tri body v auguste a septembri 2002 boli odhadnuté pomocou WBS s $C = 1$, konkrétne 12.08.2002 a 30.09.2002. V tomto období sa tiež udiali jedny z najväčších prepádov cien akcií na celosvetovom akciovom trhu. Na Obrázku 4.4 môžeme vidieť, že po udalostiach v septembri 2001 výnosy chvíľu klesali, následne je zaznamenaný rastúci trend až do roku 2002, kedy nastal prepád cien akcií.

V októbri 2002 boli pomocou WBS s $C = 1$ detekované ďalšie dva body zmeny súvisiace s globálnym prepadom cien akcií v roku 2002. Bod zmeny bol odhadnutý na 14.10.2002. V období finančnej krízy v rokoch 2007-2008 bolo celkovo detekovaných 30 bodov zmeny. Odhadnuté body zmeny pripadajú na 15.09.2008, 20.10.2008, 27.10.2008 (bol detekovaný iba pomocou WBS $C = 1,3$ pre pôvodný rad a WBS $C = 1$ pre GARCH), 02.02.2009 (bol detekovaný iba pomocou WBS $C = 1$ pre GARCH model), 02.03.2009 a 16.03.2009. Nakoniec v júli 2011 bol odhadnutý bod zmeny spadajúci na dátum 25.07.2011. Celkovo bol detekovaný šiestimi z celkových osem skúmaných metód. Domnievame sa, že súvisí s prepadom cien na trhu akcií v auguste toho istého roku, ktorý je viditeľný na Obrázku 4.4 kumulatívnych súčtov týždenných logaritmickejých výnosov.

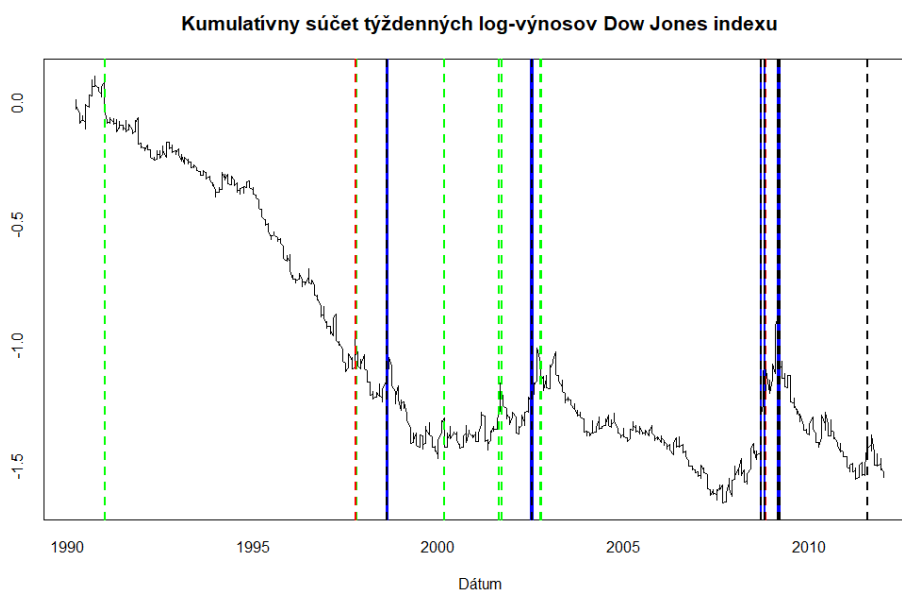
Na Obrázku 4.6 môžeme ďalej vidieť odhad strednej hodnoty týždenných logaritmickejých výnosov získaný pomocou WBS s automatickým nastavením parametrov z GARCH modelu. Nakoniec, pre porovnanie, jedna z metód, ktoré popisujú James a Matteson (2014), detekovala na danom dátovom súbore 4 body zmeny, ktoré prislúchajú dátumom 21.10.1996, 31.03.2003, 15.10.2007 a 09.03.2009.



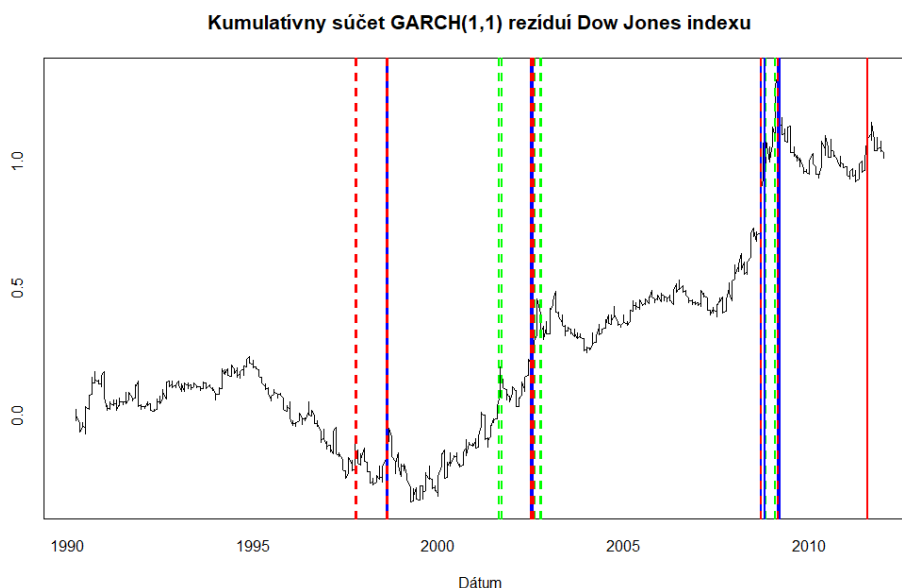
Obr. 4.2: WBS odhady polôh bodov zmeny (v mesiacoch) pre rad týždenných logaritmických výnosov Dow Jones indexu – WBS $C = 1$ (všetky vertikálne čiary); WBS $C = 1.3$ (červené, modré a čierne čiary); WBS sSIC $\alpha = 1$ (modré plné čiary); WBS auto (čierne prerušované čiary).



Obr. 4.3: WBS odhady polôh bodov zmeny (v mesiacoch) pre rezíduá týždenných logaritmických výnosov Dow Jones indexu odhadnutých pomocou GARCH(1,1) modelu s normálnymi inováciami – WBS $C = 1$ (všetky vertikálne čiary); WBS $C = 1.3$ (červené vertikálne čiary); WBS sSIC $\alpha = 1$ (modré a červené plné čiary); WBS auto (modré plné čiary).



Obr. 4.4: Kumulatívne súčty týždenných logaritmickej výnosov Dow Jones indexu s WBS odhadmi bodov zmeny – WBS $C = 1$ (všetky vertikálne čiary); WBS $C = 1.3$ (červené, modré a čierne čiary); WBS sSIC $\alpha = 1$ (modré plné čiary); WBS auto (čierne prerušované čiary).



Obr. 4.5: Kumulatívne súčty rezíduí týždenných logaritmickej výnosov Dow Jones indexu odhadnutých pomocou GARCH(1,1) modelu s normálnymi inováciami – WBS $C = 1$ (všetky vertikálne čiary); WBS $C = 1.3$ (červené vertikálne čiary čiary); WBS sSIC $\alpha = 1$ (modré a červené plné čiary); WBS auto (modré plné čiary).

Metóda	Model	jan. 1991	okt. 1997	aug. 1998
WBS sSIC $\alpha = 1$	DJI/GARCH	0	0	2
WBS auto		0	0	1
WBS C = 1,0	DJI	1	2	2
WBS C = 1,3		0	1	2
WBS C = 1,0	GARCH	0	1	2
WBS C = 1,3		0	2	2
Σ		1	6	14

Tabuľka 4.4: Počet odhadnutých bodov zmeny v jednotlivých mesiacoch pre rôzne metódy. Metódy WBS sSIC a WBS auto detekovali v daných mesiacoch pre oba modely rovnaké počty bodov zmeny.

Metóda	Model	mar. 2000	aug.2001	sep.2001
WBS sSIC $\alpha = 1$	DJI/GARCH	0	0	0
WBS auto		0	0	0
WBS C = 1,0	DJI	1	1	1
WBS C = 1,3		0	0	0
WBS C = 1,0	GARCH	0	0	0
WBS C = 1,3		0	1	1
Σ		1	2	2

Tabuľka 4.5: Počet odhadnutých bodov zmeny v jednotlivých mesiacoch pre rôzne metódy. Metódy WBS sSIC a WBS auto detekovali v daných mesiacoch pre oba modely rovnaké počty bodov zmeny.

Metóda	Model	júl 2002	aug.2002	sep.2002
WBS sSIC $\alpha = 1$	DJI/GARCH	3	0	0
WBS auto		1	0	0
WBS C = 1,0	DJI	3	0	1
WBS C = 1,3		3	0	0
WBS C = 1,0	GARCH	3	1	1
WBS C = 1,3		4	0	0
Σ		21	1	2

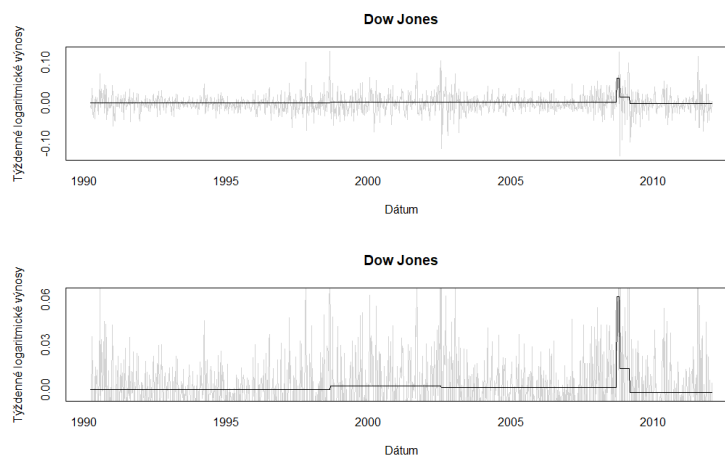
Tabuľka 4.6: Počet odhadnutých bodov zmeny v jednotlivých mesiacoch pre rôzne metódy. Metódy WBS sSIC a WBS auto detekovali v daných mesiacoch pre oba modely rovnaké počty a polohy bodov zmeny.

Metóda	Model	okt. 2002	sep.2008	okt.2008
WBS sSIC $\alpha = 1$	DJI/GARCH	0	1	1
WBS auto		0	1	1
WBS $C = 1,0$	DJI	1	1	1
WBS $C = 1,3$		0	1	2
WBS $C = 1,0$	GARCH	1	1	2
WBS $C = 1,3$		0	1	0
Σ		2	8	9

Tabuľka 4.7: Počet odhadnutých bodov zmeny v jednotlivých mesiacoch pre rôzne metódy. Metódy WBS sSIC a WBS auto detekovali v daných mesiacoch pre oba modely rovnaké počty a polohy bodov zmeny.

Metóda	Model	feb.2009	mar. 2009	júl 2011
WBS sSIC $\alpha = 1$	DJI	0	2	0
WBS $C = 1,0$		0	1	1
WBS $C = 1,3$		0	2	1
WBS auto		0	2	1
WBS sSIC $\alpha = 1$	GARCH	0	1	1
WBS $C = 1,0$		1	2	1
WBS $C = 1,3$		0	1	1
WBS auto		0	1	0
Σ		1	12	6

Tabuľka 4.8: Počet odhadnutých bodov zmeny v jednotlivých mesiacoch pre rôzne metódy.



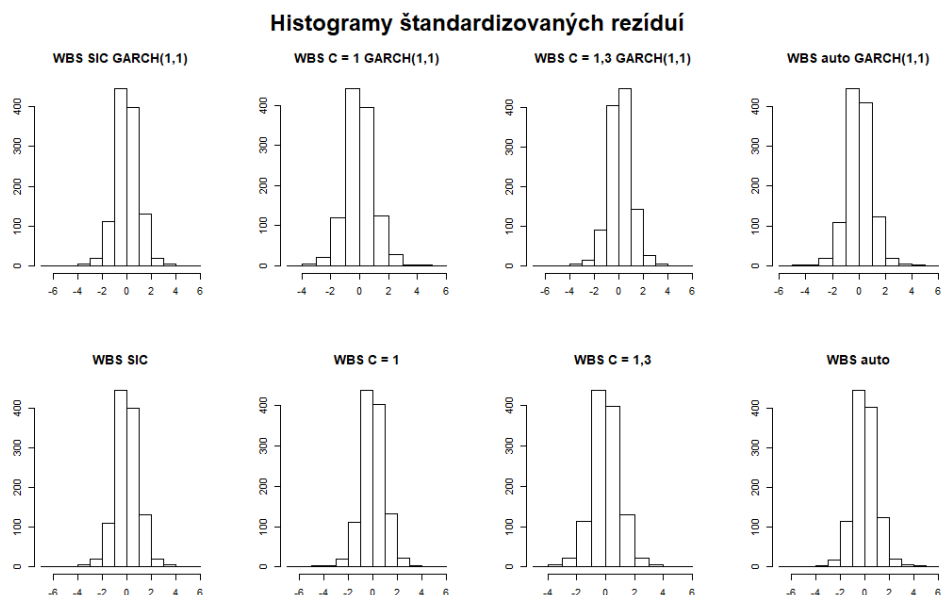
Obr. 4.6: Týždenné logaritmické výnosy Dow Jones indexu (sivá) a WBS odhad ich strednej hodnoty (čierna) s automatickým nastavením parametrov C a M pre model GARCH.

4.4 Reziduálna diagnostika

V tejto časti sa zameriame na overovanie predpokladu

$$\{\varepsilon_t\}_t \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2).$$

WBS metóda dáva najlepšie výsledky, ak dáta tento predpoklad neporušujú. Samozrejme finančné dáta často vykazujú nenormalitu, či autokoreláciu rezíduí v čase. Už z histogramov odhadnutých modelov na Obrázku 4.7 je zjavné, že odhadnuté reziduá majú zvýšenú špičatost. Hypotézu normality rezíduí testujeme pomocou Jarque-Bera testu. Výsledky sú zhrnuté v Tabuľke 4.9 a koeficienty špičatosti pre jednotlivé modely sú v Tabuľke 4.10. Pri všetkých modeloch hypotézu normality zamietame s p-hodnotou $p < 0,001$. Najnižšia hodnota Jarque-Bera testovej štatistiky je však dosiahnutá pre model GARCH s použitím WBS $C = 1$ a ďalej pre model DJI s použitím WBS $C = 1,3$. Tieto modely dosahujú tiež najnižšie hodnoty špičatosti a môžeme teda tvrdiť, že predpoklad normality porušujú zo skúmaných modelov najmenej.



Obr. 4.7: Histogramy štandardizovaných rezíduí týždenných logaritmickej výnosov Dow Jones indexu pre pôvodný rad a GARCH model a pre WBS metódy s rôznymi penalizačnými kritériami.

Ďalej budeme overovať, či sú reziduá nekorelované v čase, aj keď zrejme nemôžeme predpokladať, že týždenné výnosy akciového indexu sú nezávislé. Autokorelačné funkcie pre odhady štandardizovaných rezíduí pre jednotlivé metódy môžeme vidieť na Obrázku 4.8. Korelácie rádov vyšších než jedna sú nevýznamné. Korelácie prvého rádu pre jednotlivé metódy sú zhrnuté v Tabuľke 4.10 a príliš sa od nuly neodchyľujú. Takisto hodnoty Durbin-Watsonových testových štatistík zhrnuté v Tabuľke 4.9 naznačujú, že zrejme môžeme predpokladať nekorelovanosť rezíduí.

Metóda	Model	Test	Test. št.
WBS sSIC $\alpha = 1$	DJI	Jarque–Bera	417,47
WBS C = 1,0			442,37
WBS C = 1,3			209,55
WBS auto			653,76
WBS sSIC $\alpha = 1$	GARCH		417,47
WBS C = 1,0			189,47
WBS C = 1,3			591,49
WBS auto			680,00
Pôvodné rezíduá			660,44
WBS sSIC $\alpha = 1$	DJI	Durbin–Watson	2,23
WBS C = 1,0			2,25
WBS C = 1,3			2,19
WBS auto			2,26
WBS sSIC $\alpha = 1$	GARCH		2,23
WBS C = 1,0			2,27
WBS C = 1,3			2,17
WBS auto			2,25
Pôvodné rezíduá			2,21

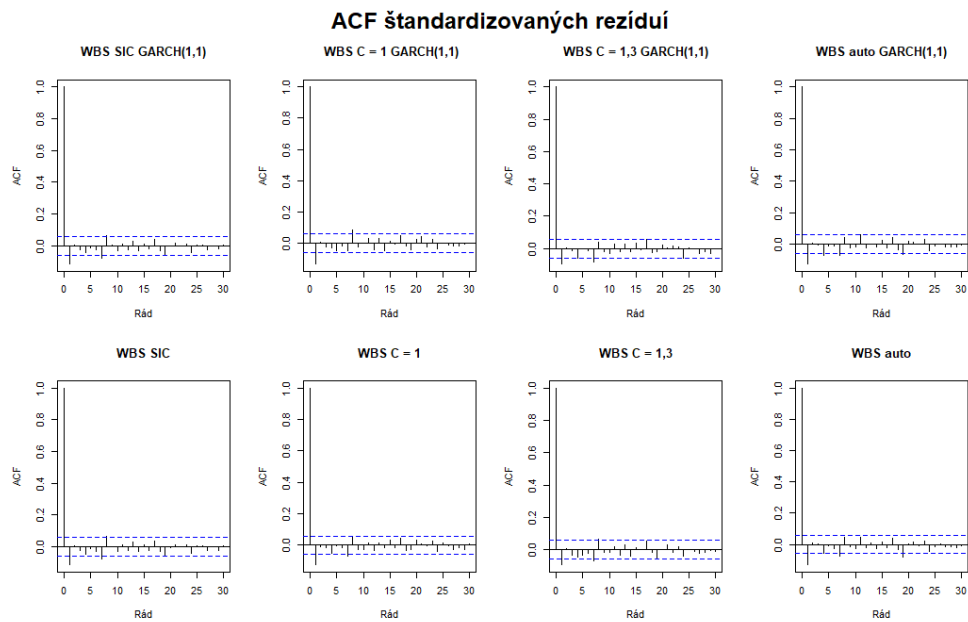
Tabuľka 4.9: Hodnoty testových štatistík pre relevantné testy na štandardizovaných rezíduách pre jednotlivé modely. Normalitu zamietame pri všetkých modeloch s $p < 0,001$.

Metóda	Model	Koeficient	Hodnota
WBS sSIC $\alpha = 1$	DJI	γ_2	2,96
WBS C = 1,0			3,01
WBS C = 1,3			2,08
WBS auto			3,69
WBS sSIC $\alpha = 1$	GARCH		2,96
WBS C = 1,0			1,94
WBS C = 1,3			3,45
WBS auto			3,77
Pôvodné rezíduá			3,67
WBS sSIC $\alpha = 1$	DJI/GARCH	ρ_1	-0,11
WBS C = 1,0			-0,13
WBS C = 1,3			-0,09
WBS auto			-0,13
Pôvodné rezíduá			-0,11

Pozn: γ_2 – Koeficient špičatosti;

ρ_1 – Koeficient autokorelácie prvého rádu.

Tabuľka 4.10: Odhadnuté hodnoty koeficientov špičatosti a korelácie pre štandardizované rezíduá z jednotlivých modelov.



Obr. 4.8: Autokorelačné funkcie štandardizovaných rezíduí týždenných logaritmickejých výnosov Dow Jones indexu pre pôvodný rad a GARCH model a pre WBS metódy s rôznymi penalizačnými kritériami.

Aj keď má použitý dátový súbor špicatejšie rozdelenie a nemôžeme zaručiť ani nezávislosť rezíduí, výsledky zhrnuté v Sekcii 4.3 sú dobre interpretovateľné. WBS metóda pomerne vierohodne detekuje zmeny súvisiace s významnými udalosťami na finančnom trhu.

Záver

Táto práca sa zaoberala odhadovaním štrukturálnych zmien v strednej hodnote časových radov. Vysvetlených bolo niekoľko prístupov k odhadovaniu počtu a polôh bodov zmeny v jednorozmerných časových radoch a pre panelové dáta. Prvé tri kapitoly boli venované teoretickému výkladu skúmaných metód. Štvrtá kapitola sa zameriavala na odhadovanie bodov zmeny v reálnom dátovom súbore.

V prvej kapitole sme zaviedli modely časového radu s jedným a viacerými bodmi zmeny v nepodmienenej strednej hodnote. Vysvetlili sme test o neprítomnosti bodu zmeny založený na preškálovanom CUSUM procese a test pomerom vierohodností. Nakoniec sme stručne popísali metódu odhadovania počtu bodov zmeny pomocou Schwarzovho kritéria.

Druhá kapitola bola zameraná na výklad metódy WBS, ktorá vychádza z binárnej segmentácie a je založená na použití náhodných intervalov. Zavedené boli predpoklady pre fungovanie binárnej segmentácie a metódy WBS. Ďalej boli popísané algoritmy fungovania týchto metód a podmienky, za ktorých sú odhady bodov zmeny konzistentné. V závere kapitoly bolo popísané zosilnené Schwarzovo informačné kritérium a podmienky pre konzistenciu odhadov získaných s jeho použitím.

Tretia kapitola sa venovala odhadovaniu bodov zmeny v panelových dátach. Modely zavedené v Kapitole 1 sme rozšírili pre prípad viacerých časových radov. Vysvetlené boli metódy pre pevný a pre veľký počet pozorovaní, podmienky pre ich konzistenciu a pre konzistenciu zavedených testových štatistík. Boli popísané dva prístupy k modelovaniu závislosti medzi panelmi.

V štvrtej kapitole sme analyzovali týždenné logaritmické výnosy Dow Jones akciového indexu napočítané zo záverečných kurzov. Na tomto dátovom súbore sme odhadovali počet a polohy bodov zmeny pomocou metódy WBS s rôznym nastavením parametrov a rôznymi penalizačnými kritériami. Pôvodný rad sme najskôr očistili od zhlukov volatility aplikáciou GARCH(1,1) modelu so štandardnými normálnymi inováciami. Metódu WBS sme použili na pôvodný rad výnosov a na odhadnuté rezíduá z modelu GARCH. Odhadnuté rezíduá modelov s bodmi zmeny mali značne zvýšenú špičatosť. Rezíduá získané pomocou WBS s ručnou voľbou parametrov porušovali predpoklad normality najmenej. Autokorelácia vyšších rádo v dátach nebola prítomná a autokorelácie prvého rádu boli zanedbateľne malé. Nezávislosť sme ale predpokladať nemohli. WBS metóda napriek porušeniu predpokladu normality v dátovom súbore dávala uspokojivé a dobre interpretovateľné výsledky. Odhadnuté polohy bodov zmeny spadali do období, kedy bol ich výskyt očakávaný. Súviseli najmä s významnými udalosťami na finančnom trhu, ako bola hospodárska kríza v rokoch 2007 až 2009. Priemerne najvyšší počet bodov zmeny odhadovali metódy s ručnou voľbou parametrov, najmenej ich odhadla metóda s automatickou voľbou parametrov. Prvé dve spomínané metódy tiež odhadovali body zmeny frekventovanejšie, s menšími vzdialenosťami, pričom často odhadli viac než jeden bod zmeny v priebehu jedného mesiaca. Naopak WBS s automatickou voľbou parametrov odhadla nanajvyš jeden bod zmeny za mesiac.

Zoznam použitej literatúry

- ANDREWS, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, **59**(3), 817–858.
- AUE, A. a HORVÁTH, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, **34**(1), 1 – 16.
- BAI, J. (2010). Common breaks in means and variances for panel data. *Journal of Econometrics*, **157**(1), 78 – 92.
- BARANOWSKI, R., CHEN, Y. a FRYZLEWICZ, P. (2019). Narrowest-over-threshold detection of multiple change-points and change-point-like features. *Journal of the Royal Statistical Society Series B*, **81**, 649–672.
- CIPRA, T. (2008). *Finanční ekonometrie*. 1. vydání. Ekopress, Praha. ISBN 978-80-86929-43-9.
- DOUKHAN, P. (1994). *Mixing: Properties and Examples*. Lecture Notes in Statistics (Book 85). Springer-Verlag New York Inc., New York, NY, United States. ISBN 978-1-4612-2642-0.
- FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, **42**(6), 2243 – 2281.
- FRYZLEWICZ, P. (2017). Package 'breakfast' : Multiple change-point detection and segmentation. <https://cran.r-project.org/web/packages/breakfast/breakfast.pdf>. Prístup: 2020-05-15.
- HORVÁTH, L. a HUŠKOVÁ, M. (2012). Change-point detection in panel data. *Journal of Time Series Analysis*, **33**(4), 631 – 648.
- JAMES, N. A. a MATTESON, D. S. (2014). ecp: An r package for nonparametric multiple change point analysis. *Journal of Statistical Software*, **62**(7), 1–25.
- MACIAK, M., PEŠTA, M. a PEŠTOVÁ, B. (2020). Changepoint in dependent and non-stationary panels. *Statistical Papers*. doi: 10.1007/s00362-020-01180-6.
- MICHEAS, A. C. (2018). *Theory of Stochastic Objects: Probability, Stochastic Processes and Inference*. Taylor and Francis Group, LLC, Boca Raton, Florida, USA. ISBN 978-1-4665-1520-8.
- PEŠTA, M., PEŠTOVÁ, B. a MACIAK, M. (2020). Changepoint estimation for dependent and non-stationary panels. *Applications of Mathematics*. doi: 10.21136/AM.2020.0296-19.
- PEŠTOVÁ, B. a PEŠTA, M. (2017). Change point estimation in panel data without boundary issue. *Risks*, **5**(1), 7.
- PRÁŠKOVÁ, Z. (2004). *Základy náhodných procesů II*. 1. vydání. Karolinum, Praha. ISBN 978-80-246-3516-3.

- WUERTZ, D. (2020). Package 'fgarch': Rmetrics - autoregressive conditional heteroskedastic modelling. <https://cran.r-project.org/web/packages/fGarch/fGarch.pdf>. Pristup: 2020-05-15.
- YAO, Y.-C. (1988). Estimating the number of change-points via schwarz' criterion. *Statistics and Probability Letters*, **6**, 181–9.

Zoznam obrázkov

4.1	Týždenné logaritmické výnosy Dow Jones indexu a volatilita odhadnutá pomocou GARCH(1,1) modelu s normálnymi inováciami.	31
4.2	WBS odhady polôh bodov zmeny (v mesiacoch) pre rad týždenných logaritmických výnosov Dow Jones indexu – WBS $C = 1$ (všetky vertikálne čiary); WBS $C = 1.3$ (červené, modré a čierne čiary); WBS sSIC $\alpha = 1$ (modré plné čiary); WBS auto (čierne prerušované čiary).	35
4.3	WBS odhady polôh bodov zmeny (v mesiacoch) pre rezíduá týždenných logaritmických výnosov Dow Jones indexu odhadnutých pomocou GARCH(1,1) modelu s normálnymi inováciami – WBS $C = 1$ (všetky vertikálne čiary); WBS $C = 1.3$ (červené vertikálne čiary); WBS sSIC $\alpha = 1$ (modré a červené plné čiary); WBS auto (modré plné čiary).	35
4.4	Kumulatívne súčty týždenných logaritmických výnosov Dow Jones indexu s WBS odhadmi bodov zmeny – WBS $C = 1$ (všetky vertikálne čiary); WBS $C = 1.3$ (červené, modré a čierne čiary); WBS sSIC $\alpha = 1$ (modré plné čiary); WBS auto (čierne prerušované čiary).	36
4.5	Kumulatívne súčty rezíduí týždenných logaritmických výnosov Dow Jones indexu odhadnutých pomocou GARCH(1,1) modelu s normálnymi inováciami – WBS $C = 1$ (všetky vertikálne čiary); WBS $C = 1.3$ (červené vertikálne čiary); WBS sSIC $\alpha = 1$ (modré a červené plné čiary); WBS auto (modré plné čiary).	36
4.6	Týždenné logaritmické výnosy Dow Jones indexu (sivá) a WBS odhad ich strednej hodnoty (čierna) s automatickým nastavením parametrov C a M pre model GARCH.	38
4.7	Histogramy štandardizovaných rezíduí týždenných logaritmických výnosov Dow Jones indexu pre pôvodný rad a GARCH model a pre WBS metódy s rôznymi penalizačnými kritériami.	39
4.8	Autokorelačné funkcie štandardizovaných rezíduí týždenných logaritmických výnosov Dow Jones indexu pre pôvodný rad a GARCH model a pre WBS metódy s rôznymi penalizačnými kritériami.	41

Zoznam tabuliek

4.1	Odhady koeficientov GARCH(1,1) modelu so štandardnými normálnymi inováciami, odhady štandardných odchýlok a hodnoty testových štatistík t-testov s príslušnými p-hodnotami.	31
4.2	Popisné charakteristiky odhadnutého počtu bodov zmeny \hat{m} napočítané zo 100 prevedení metód.	33
4.3	Popisné charakteristiky odhadnutého počtu bodov zmeny \hat{m} napočítané zo 100 prevedení metód.	33
4.4	Počet odhadnutých bodov zmeny v jednotlivých mesiacoch pre rôzne metódy. Metódy WBS sSIC a WBS auto detekovali v daných mesiacoch pre oba modely rovnaké počty bodov zmeny. . . .	37
4.5	Počet odhadnutých bodov zmeny v jednotlivých mesiacoch pre rôzne metódy. Metódy WBS sSIC a WBS auto detekovali v daných mesiacoch pre oba modely rovnaké počty bodov zmeny. . . .	37
4.6	Počet odhadnutých bodov zmeny v jednotlivých mesiacoch pre rôzne metódy. Metódy WBS sSIC a WBS auto detekovali v daných mesiacoch pre oba modely rovnaké počty a polohy bodov zmeny.	37
4.7	Počet odhadnutých bodov zmeny v jednotlivých mesiacoch pre rôzne metódy. Metódy WBS sSIC a WBS auto detekovali v daných mesiacoch pre oba modely rovnaké počty a polohy bodov zmeny.	38
4.8	Počet odhadnutých bodov zmeny v jednotlivých mesiacoch pre rôzne metódy.	38
4.9	Hodnoty testových štatistík pre relevantné testy na štandardizovaných rezíduách pre jednotlivé modely. Normalitu zamietame pri všetkých modeloch s $p < 0,001$	40
4.10	Odhadnuté hodnoty koeficientov špičatosti a korelácie pre štandardizované rezíduá z jednotlivých modelov.	40

Zoznam použitých skratiek

$\xrightarrow[T \rightarrow \infty]{\mathcal{D}}$ konvergencia v distribúcii

i.i.d. nezávislé rovnako rozdelené náhodné veličiny

[.] dolná celá časť čísla

[.] horná celá časť čísla

$\text{Cor}(X, Y)$ korelácia X a Y

$\text{Cov}(X, Y)$ kovariancia X a Y